Hybrid flexible (HyFlex) seminar delivery – A technical overview of the implementation

Rodrigo Sanchez-Pizani [a],[*], Michael Detyna [b], Stephen Dance [c], Luis Gomez-Agustina [c]

[a] *London South Bank University/King's College London, United Kingdom*
[b] *King's College London, United Kingdom*
[c] *London South Bank University, United Kingdom*

ABSTRACT

This paper investigates a new technology for Hybrid flexible delivery (known as HyFlex), as implemented at King's College London. The relatively novel character of HyFlex, of mixing synchronously on-line and in-room teaching, and the recent changes due to the COVID-19 pandemic mean this use of the technology and teaching model is largely new to the UK. This research evaluated audio quality in the context of a HyFlex technical environment. The paper provides a high-level overview of the process of designing a HyFlex solution and presents a detailed evaluation of the impact of reverberation in relation to the accuracy of automatically generated subtitles and the influence of microphone selection. The paper shows that there was a significant relationship between the reverberation, the audio quality, and the subtitling system, which is important as past studies highlighted audio quality is key for the students' experience. It presents a viable and simple methodology to estimate the audio quality on installed HyFlex systems to improve the students experience in a hybrid teaching environment.

1. Introduction

The present paper describes the rationalisation behind the design of King's College London (KCL) HyFlex seminar rooms and provides an overview of possible mechanisms that can be used to enhance the hybrid educational environment model.

This paper focuses on the technology and design behind the implementation of HyFlex spaces as deployed in several classroom environments at King's College London. The research aims to resolve the challenge of students, some of whom wish to attend in person and others on-line. The delivery mechanism discussed and evaluated here is commonly referred in the consulted literature as HyFlex, and in the case of this research it focuses only on the delivery of seminars in such a way that the student can join the event on campus or on-line synchronously. HyFlex involves the lecturer and some students being on campus, whilst other students join the event via a web meeting with cameras, microphones and screens connecting both groups simultaneously. King's adopted this approach for several faculties.

The current pandemic resulting from COVID-19 has had a global impact [1], and has led to a move away from face-to-face teaching and a move towards on-line teaching [2]; thus, forcing the hand of universities to implement more creative ways to teach while maintaining high quality learning interaction. A comprehensible systematic literature review on the subject, by Ref. [3]provided an overview of the benefits,

challenges and design principles of synchronous hybrid learning, based on 47 modern research studies. They argued that compared to fully on-line or fully on site, synchronous hybrid is a more flexible and more engaging learning space.

The paper is structured as follows: a literature review of the main concepts to be discussed in the manuscript, followed by the methodology of evaluation of some of the systems in use and a description of the implementation, plus a discussion of the main findings and conclusions.

2. Literature

The hybrid flexible approach of having seminars delivering synchronously face-to-face and on-line lessons (HyFlex) was novel to UK institutions. Approaches similar to HyFlex have been adopted in an Australian university under the name "Dual mode" [4].

For the remainder of this paper, **HyFlex** means fully synchronous teaching with both cohorts interacting.

2.1. HyFlex

An example of early set-ups for HyFlex sessions was undertaken by Ref. [5] with teachers in remote locations using video conferencing; the authors concluded that the use of the video conference technology allowed participants good interaction whilst keeping contact with their

---

remote location.

The importance of general interaction between students and teachers with the technology used in teaching spaces has also been addressed in the past by Ref. [6] where they discuss the benefits of a flipped classroom (pedagogical approach where students read or view the material first and then discuss or conduct activities in class) in a South African higher education institution. Although these were not generalisable due to limitations of the study, it showed that there is an influence between the understanding of the technology and outcomes.

[7] discuss a simple methodology to evaluate the introduction of new technologies for teaching by using student's response systems (SRSs) and the value of testing the technology on small focus groups; however, this is limited to the interaction with the technology.

The on-line book Expanding the Implementation of Hybrid-Flexible Courses and Programs by Ref. [8]; includes details concerning HyFlex and how to make the most of it. Amongst the recommendations highlighted is to "work with the techies"; however, there seems to be no reference to the actual performance of the space, the technology used and its relationship with the success of HyFlex implementations.

It appears that a comparison between objective metrics and subjective responses is lacking. A systematic literature review by Ref. [3] seems to substantiate this argument, that there is little research and evaluation of Hybrid set-ups that consider objective technical measurable parameters and cognitive parameters simultaneously.

The work from Ref. [9]tries to understand the implications for web based technologies, like lecture capture, and the implications when used in learning and teaching environments; some of the findings are later replicated by Ref. [10]. While [11] focuses on investigating the room design.

The research presented by Ref. [12] briefly mentions the technology used; however, no objective metrics can be identified as the overall research focused on areas like teaching presence, social presence, cognitive presence, learners' presence and not the technology used.

2.2. Objective parameters

There has been a substantial amount of research that evaluates certain elements of technology or room parameters objectively; but these are either several years old, e.g. Ref. [13], used 37 volunteers to compare acoustical evaluation of speakers on site noting the importance of replicating listeners conditions. Or concentrate on test scenarios and not on production environments, e.g., Relation to subjective assessment of the quality of the end-to-end signal. J. G [14]. worked on a development the Perceptual Objective Listening Quality Assessment (POLQA) an end-to-end objective audio quality assessment method that has been adopted, as an standard methodology. This has now been incorporated into ITU-T P.863 [15].

A recommendation for the subjective evaluation of speech quality was published by the International Telecommunication Union (ITU) in order to analyse the quality of codecs using Mean Opinion Scores (MOS) [50]. These are quality ratings given by listeners (not necessarily trained or experts) to the quality evaluation of a recording and provides an average quality rating of a particular recording. MOS scales can be used, in two-way communications or listening mode only.

MOS scores have then been used by other researchers e.g. Ref. [16] developed a single-ended method to quantify audio quality in the presence of nonlinear distortions using HASQI (Hearing Aid Sound Quality Index) and found that the method had an error 0.17 units in the MOS scale for Music.

The MOS scale was then extended by ITU and used by Ref. [17]in the webMUSHRA as an on-line implementation.

2.3. Related work

The recent work by Ref. [18] provided a study of the distant-talking speech recognition in reverberant conditions for simulated and real conditions. By comparing the room impulse response to the Automatic Speech Recognition (ASR) accuracy and early to late reverberation [19], compared multiple microphones set-ups and acoustical conditions; however their approach was intended to improve the performance of ASR systems and not as a method to identify quality.

In the case presented by Ref. [20] an important highlight of the use of technology and its impact on the success of teaching synchronous models was made, summarised as Here or There (HOT), i.e. allowing students on campus or on-line to interact with each other. In here, although there is a discussion about the look and feel of the technology, there is no mention of objective metrics to compare and evaluate the apparatus in use or the technological vehicles. The document does, however, include an important comment that the decision should not be guided purely from pedagogical or technological factors but as a combination of both.

2.4. Building environment and acoustics

King's College London (KCL) adopted the view that good room acoustics is paramount for in room student experience and incorporated acoustic parameters as part of the audio-visual standards document used by the university in teaching rooms [21]. study the impact of reverberation and noise in informal learning spaces (ILS) concluding that acoustics take an important role in these areas [22]. presented an analysis of six higher education classrooms to study the impact of classroom attributes on student satisfaction and performance. Amongst their findings was the difference between student perception of acoustic conditions of the rooms and the impact on their learning, making acoustics of the highest importance for a good in room experience. An examination into acoustic treatment in classrooms was investigated by Ref. [23] where they found a significant relationship between the acoustics of the room and intelligibility, which also helped the students concentration.

One of the metrics commonly used to evaluate the audio (speech) degradations of the acoustic channel is the speech transmission index (STI) introduced by Ref. [24]. [25] validate these results in real conditions and [26] proposed a reviewed spectrum of the previously mentioned index that was later incorporated in BS EN IEC 60268-16-2020 [27]. Since 2019, King's has incorporated this as an objective metric for new classrooms.

A set of performance standards for acoustic design in education have previously been written up in BB93 [28]. This highlights the importance of acoustic design in schools and establishes a series of recommendations. An equivalent document does not exist for higher education (HE) institutions in the UK. As result King's adapted some of the parameters in BB93 [29].

The literature seems scarce with regard to investigating the technology and the building environment of HyFlex; this manuscript intends to contribute to fill that gap.

3. Methodology

The methodology for the research on acoustics is detailed below, however it is important to be aware that, prior to this research, the HyFlex technical solution itself had to be implemented. It was important to produce a solution that was viable and pedagogically sound. To achieve this, a framework had to be created incorporating technical aspects and addressing concerns of a wider audience, integrating experts on the knowledge domains, involved in the decision-making process.

[30] identifies 4 stages in the knowledge acquisition process. i.e., planning, knowledge extraction, knowledge analysis and knowledge verification. Although these are proposed as a way to design expert systems, the process is transferable to technology design or other domains, as shown by Ref. [31] where the above are somewhat included as part of the proposed seven phases framework: prepare (α), discover (β), define (γ), ideate (δ), validate (ε), implement (ζ), and deliver(η).

Defining the technology selection to be used was achieved empirically as no clear methodology to select equipment was found in the literature. One of the clear trends observed in the literature was the importance of quality sound and one of the requirements of HyFlex is to replicate the "then and there" [32]). i.e., a student listening to an audio recording, or a lecture remotely should expect to have a similar listening experience to a student present in the classroom. It is reasonable to assume that it is necessary to understand the original acoustic environment to establish what is rational to expect from the far end i.e., if the in-room acoustics are poor it is unreasonable to assume that the far end will have good audio.

Spaces that had been recently refurbished were selected and acoustic of the spaces was judged subjectively by four audio experts. Similarly, feedback received over the years was used to discard rooms that have received complaints about sound quality (mostly related to reverberation) and poor visibility.

An alternative to evaluate the microphones was proposed. This consisted of using Automatic Speech Recognition (ASR) and juxtaposing the results from this with the original text. To compare the result, it was proposed to use the Levenshtein distance. The latter has been used in the past as a method to identify plagiarism [33]. When used in this way the higher the Levenshtein score, the greater the difference between the text and transcription. This research aimed to find greater similarities between texts, so the inverse has been used, i.e., the higher the score, the better the similarity.

The initial hypothesis investigated if there is a correlation between subjective appreciation scores and the automatic subtitling system. The results showed whether this hypothesis was correct. The secondary hypothesis investigated if there was a significant relationship between the microphone position in the room and the subjective scores in user perceptions. The results evaluated the evidence for this and the benefit (or lack of it) for moving the microphones around in the room.

To deliver a good design it was necessary to separate the research into several sections (α). The first part occurs prior to the design (δ) of HyFlex and seeks to find the best microphone to be used (β) the remaining investigates the subjective performance of the microphones (ε) and the implementation (ζ) and delivery (η).

3.1. HyFlex, defining the concept (α)

It is important to keep in mind that as [34] argues, simply reusing an existing curriculum may not be effective with HyFlex and may not unite physical and virtual students. This is in line with [35]; who highlighted the risk of making HyFlex all about investing in technology.

3.1.1. Initial considerations

A concern, when combining in room and online audiences, is that the former could diminish the experience of the latter [36]. However, proper classroom management, use of suitable microphone and providing students with personal devices (used for chats and online tools, no audio to avoid feedback), can help mitigate these potential challenges.

In a separate, internal investigation King's researchers discussed the implementation of HyFlex spaces and the importance of finding a balance between pedagogy and technology. Amongst the main requirements highlighted, there were three that were used to bound the problem:

- The solution must be scalable, with minimum workload on the teaching staff, encourage debate as well as student-student and teacher-student interaction replicating the "then and there" while being remote.
- From the technical perspective, the researchers highlighted the importance of good quality audio and intelligibility, as well as camera positioning.

- As a result of Covid-19, there was a necessity to wear face masks; therefore, it was considered important to evaluate the impact of face masks on the design, given lack of research found on the subject.

This paper concentrates on the first two points.

3.2. Microphones Selection(β)

The standard procedure to characterise a microphone is to use the substitution methodology as this produces more accurate results; however, this assumes a controlled acoustic environment as mentioned by BS EN 60268–4:2014 [37].

To replicate the controlled environment, the microphones were taken to the Anechoic room (Fig. 1 & Fig. 2) and reverberation lab Fig. 3 in London South Bank University[1] where their sensitivity was measured (see Table 1)[2]

A series of lists [38] were recorded using a loudspeaker of suitable electro-acoustic characteristics to mimic human speech production (Fostex, model 6301 N). Audio level of the speaker was set, when measured at 1 m by a class 1 calibrated sound level meter (NTI XL2 with a M2230 microphone), to 63 dBA for speech and 66 dBA for pink noise; to follow what was done previously by other researchers (e.g. Ref. [16]) and to align with BSI [27].

A selection of 6 microphones, pre-screened during the Integrated Systems Europe2019[3] was used to record the list. Two Earthworks M50 of known sensitivity (one on stand parallel to the floor and one perpendicular) were used for benchmarking.

Three main environments were used, with the intention of evaluating the impact of reverberation (values are presented as T30 measured on a third octave band for 1 kHz centre frequency).

- The Anechoic Chamber (T30, 1 KHz ~0s)
- The reverberation lab without any absorption (T30, 1 KHz ~5s)
- The reverberation lab using the built-in absorption (T30,1 KHz ~2.5s)

To evaluate the behaviour on quasi real-life conditions microphones were also set up in a mock classroom, at King's College London (see Fig. 4).

Nine professional audio-visual engineers, all working in Audio Visual Ser-vices at King's College London (KCL), volunteered to take part in a survey, as listeners, to evaluate the quality of the recordings (under ethical approval from LSBU: ETH1819-0086 and KCL: MRA-18/

Fig. 1. Test Microphones with Pink Noise using Substitution.
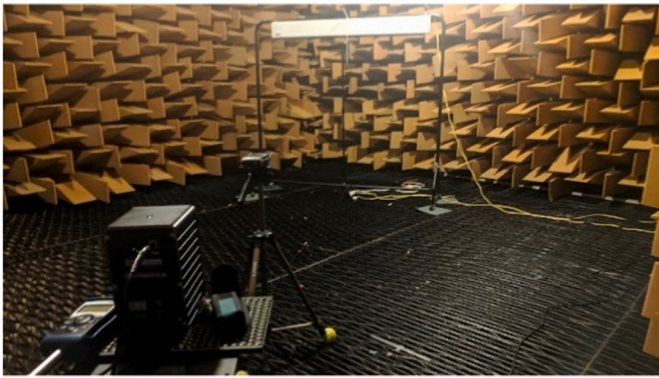

Fig. 2. Recording list anechoic room.

Fig. 3. In semi-reverberant conditions.

Table 1
Microphone Sensitivity at 1 m using a 440 Hz sinusoidal tone and.

1 m from Source, 400 Hz Tone	NTI Value	Mic result	Sensitivity
	Leq,15s(dBZ)	Leq,15s (dBZ)	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

Fig. 4. Mock classroom in position.

19–11895).

A listening station was set up. This consisted of a Dell Latitude laptop running Windows 10, a multi-track 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, this following the methodology by previous researchers e.g., Ref. [39]. For the lab recordings, the AV engineers compared each recording against the original to evaluate the degradation while, for the in-room recordings, the comparison was undertaken by comparing the recordings between each other and with the original, both procedures showed comparable results.

All the tests were evaluated using a MOS (mean opinion scores) 1–5 scale. This is to match recommendations by ITU, but it also matches with metrics used commonly by communications software where a five stars rating is given to calls; and it has been used by previous researchers, e.g. User Generated Content (UGC) [40]. Therefore, keeping a 1–5 scale allows studies to be comparable.

Three ASR (automatic speech recognition) production algorithms from Google YouTube (YT),[4] Amazon Web Services[5] (AWS) and Cielo24[6] (free version), the last two as part of King's lecture capture system (Echo360[7]), were used to compare the laboratory and in-room recordings subjective score with their overall accuracy of the text.

3.3. Video and Camera(β)

The quality of video was not something highlighted in the researched literature as one of the pressing issues for the delivery of HyFlex, instead the main video points were around how to frame the image from the camera (see for example [20].

It also considered that the outside lighting conditions would not significantly impact the light in the room nor be noticeable in the video image in the room; therefore, abiding by recommendations of [41,42].

This assumption was possible because, as part of the ongoing modernisation of the audio-visual estate, there has been a requirement for projectors and displays in rooms to be bright enough and light reflections minimised while under full artificial lighting. Having brighter displays mean that lights in the room can be in full brightness allowing more flexibility between f-number [43] shutter speeds and general camera optics. This way the only aspects to be considered were the camera connectivity, digital resolution, and control. These were defined as USB-3, 4K and AI capabilities to allow for auto framing of the image.

For video, all the decisions were made based on information supplied by manufacturers and, as the parameters were somewhat generous, the camera selection (once stratified by resolution, connectivity and auto framing capabilities) was very much dictated by size and on-line reviews.

The screen was situated on the side of the room parallel to the line of action between the lecturer and students in the room, to allow this to act as reference but not to be a distraction (see Fig. 11 in the appendix section).

For schematics of the space please see the appendixes (Fig. 12& Fig. 13) at the end of the document.

3.4. Video conferencing Technology(β)

Different communications platforms were considered for the test, this were: Zoom, Echo360, and MS Teams. But the selection was influenced by the larger institution that had adopted MS Teams as the main video conferencing vehicle.

Although the decision and evaluation of video conference technology was dictated by availability and intuitional decision, this was not

---

[4] https://www.youtube.com/
[5] https://aws.amazon.com/
[6] https://cielo24.com/
[7] https://echo360.com/

considered problematic as feedback obtained in previous years of asynchronous teaching (e.g., lecture capture) did not highlighted the need to investigate video quality.

The high-definition camera did bring a number of challenges, the higher the resolution of the cameras, the higher demand for bandwidth, T [44]. recommends 100Mbits/s for an SD signal while it recommends 400Mbit/s for HD and 1Gbit/s for 4K, and similar to this, higher resolutions will require larger displays as per [42].

To avoid the issues above and for practical reasons, a series of compromises were adopted:

- Bandwidth was limited to High Definition (HD, 1080p) quality.
- Size of the screen was limited to 85" diagonal as this would cover most of the teaching area for most scenarios as the maximum viewing distance for AVIXA'S recommendation for Basic Decision Making is 6.5 m from the screen centre.
- Cameras were limited to digital pan, tilt and zoom (PTZ). 4K Cameras were used as this would allow to zoom in with minimal loss of quality when downscaled to 1080p frames.
- To allow flexibility, the secondary screen was mounted on a trolley to allow for pivot and slight readjustment for different teaching situations

4. Results

All three algorithms from Google, Amazon and Cielo24, were fed with the recordings and the text compared with the original text from Ref. [38] The comparison Levenshtein score (LS) [33] was calculated using an on-line tool,[8] and then compared with the inverse normalised LS given by the equation below:

$$N_{LS} = \frac{5}{4\left(\frac{x-s}{b-s}\right)+1} \tag{1}$$

Where, x is the un-normalised (LS), s and b are the minimum and maximum value of the data set respectively. The results can be observed in Table 2 showing the correlation between vectors of values.

One notable result, in the lab, was the large effect that reverberation had on the accuracy of the subtitling system. This can be appreciated in Fig. 5. Where, as the reverberation time (horizontal) increases, the average accuracy score (vertical) decreases.

This finding is key for student satisfaction, in that it is known from existing literature [45] that lower acoustic quality significantly impacts student satisfaction. As longer reverberation times reduce the quality of audio (as measured by accuracy of subtitling), this reduces audio quality. The results suggest that, to improve student learning quality, attention should be paid to reduce reverberation.

The study was limited to lab conditions and used extreme conditions of reverberation. Care must be taken before deriving conclusions other than hypothesising that reverberation will have a negative impact in the accuracy of ASRs. Further study is required to understand this relationship.

Table 2

Similarity matrix in lab conditions between ASR algorithms from NG (Google), NA (Amazon), NC(Cielo24) and mean subjective (MOS) score by experts.

	NG	NA	NC	Subjective
NG	1.00	.94	.87	.80
NA	.94	1.00	.96	.83
NC	.87	.96	1.00	.79
Subjective	.80	.83	.79	1.00

A similar process was followed using real conditions in a test room, at King's College London (KCL). Microphones were set up in typical positions and connected to a Focusrite Scarlett 18i20 sound card. As described in Fig. 5.

In total 49 speaker position and microphone combinations were recorded and once again evaluated by the same group of nine audio specialists. Results from the Pearson correlations between the variables can be appreciated in Table 3.

Pearson Correlation between Vectors of Values.

When averages were compared, the Pearson correlation of the mean between the accuracy of all three ASR systems and the average subjective score was 0.83.

Table 4 shows ANOVA of the ASR and subjective scores for different source microphone position combination in the test classroom shown in Fig. 6.

The complete comparison between microphones, can be seen in Fig. 7. It shows the correlation between the subjective scores and the normalised Levenshtein score when applied to the ASR algorithms. With $r^2 = 0.691$ and 95% confidence level for individual values.

It is possible to see that the ASR results will underestimate the subjective score with the empirical relationship:

$$SS = -0.24 + 1.17(N_{LS}) \tag{2}$$

where SS is the Subjective Score and NLS the results from (1).

Fig. 8 on the other hand shows a microphone comparison stratified by subjective score and average performance of subtitling system (including Cielo24).

The results from the microphone evaluation were combined into a decision matrix, Table 5, that contained additionally ease of installation, cost, and portability. See Table 6 for details of microphone type and installation.

The last two microphones in Table 5 were not considered in the overall microphone comparison Fig. 8 due to combination of availability during testing, complexity of installation and overall performance of the microphones, resulted in too few measurements points available for microphones 7 & 8. All 8 microphones results can be observed in Fig. 9 & Fig. 10 in the annex section.

It is important to note two considerations at this stage. As the HyFlex design discussed in this manuscript investigates its use in seminar situations with an unattended setup (i.e., no technician present during class). It was paramount that any choice of microphones could capture the interaction of the whole class, be easy to deploy and required minimal human manipulation. Therefore, alternatives with individual microphones (e.g., push to talk) or where only the presenter wore a microphone (e.g., clip on microphone) are not considered as these are either too difficult to deploy or would not capture the whole classroom interaction. The use of microphones that are not in proximity or in contact with the users (e.g., table or handheld microphones) would also help to minimise manipulation noise and other more impulsive noise as for example paper ruffling and typing. The trade-off is that, by being further apart the background noise would increase; however, it was thought that noise cancelling used by the web conferencing software would help to minimise this.

As mentioned in the previous sections, KCL adopted a hybrid methodology of teaching as a response to Covid-19 and the "new normal".

Once this matrix was evaluated by King's AV Solutions team, it was concluded that the best alternatives to equip the HyFlex rooms were threefold, as shown in Table 7 (γ).

It was considered that if this is going to be a new way of teaching, it is important to incorporate these findings into any new HyFlex rooms.

5. Discussion

Overall, the results obtained from the detailed microphone evaluations (and more superficial analysis of the video and transport) drove
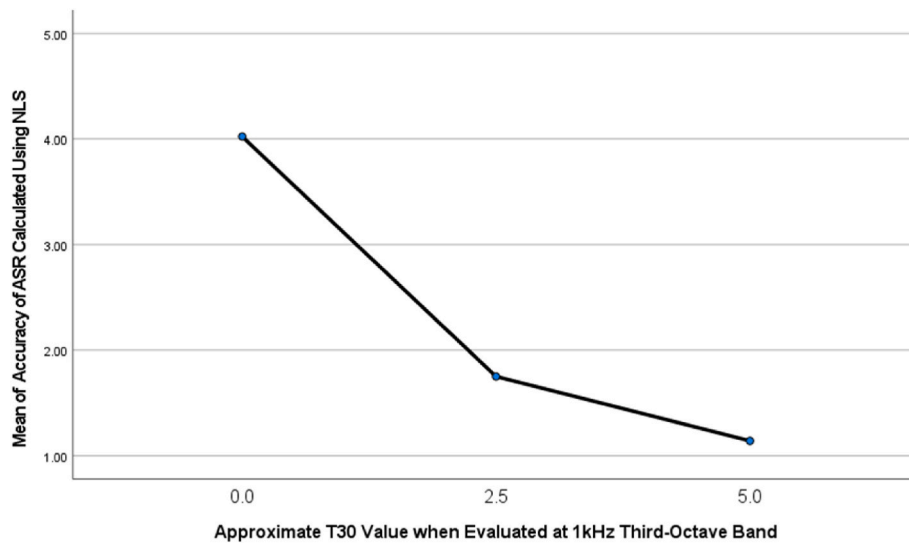
Fig. 5. Average Accuracy of all three ASR under different conditions of reverberation.

Table 3
Similarity Matrix Real Room between ASR algorithms from (Cielo24),(Google), AWS, and mean subjective (MOS) score by experts.

	Alpha Cielo24	Google (YouTube)	AWS	Subjective Score
Alpha Cielo24	1.00	0.71	0.69	0.66
Google (you tube)	0.74	1.00	0.83	0.79
AWS	0.69	0.83	1.00	0.80
Subjective Score	0.66	0.79	0.80	1.00

Table 4
ANOVA subjective scores and ASR with different speaker position.

		Sum of Squares	df	Mean Square	F	Sig.
	Between Groups	0.27	6	0.05	0.07	0.998
Subjective	Within Groups	26.23	42	0.62		
	Total	26.50	48			
	Between Groups	3.69	6	0.62	0.61	0.719
ASR	Within Groups	42.22	42	1.01		
	Total	45.91	48			

decision-making. These were then used during the overall pedagogical evaluation of the system, in the form of mock sessions and subjective evaluation of the test system.

Given the close relationship found in the literature and by the researchers, between audio quality and student experience it is evident that the selection of microphones required a more in-depth analysis and research.

The initial hypothesis, to investigate if there was a correlation between subjective appreciation scores and the automatic subtitling system appears to hold as there was good correlation between the subjective evaluation of the audio quality using MOS scale and the number of words correctly interpreted by the commercial subtitling system ($r^2 \approx 0.7$).

The secondary hypothesis that, there is a significant relationship between the microphone position in the room and the subjective score, seems unclear as there was not any significant difference; however, this might be a result of the size of the sample and as such this result should

not be considered conclusive.

The relative ease of the process could provide a cheap and quick way for institutions to decide between systems when testing these on location.

The process described in the previous subsections allowed the team to determine (γ) the equipment and system to be used (δ)

One of the key points, and an important consideration in designing HyFlex, was thinking in terms of how the students feel and how it is presented as an alternative to other forms of teaching. It was important during this research to ensure the students feel part of the experience which means a high-quality system, with high quality microphones, etc. This helped to ensure there was a good relationship between on-line and on-campus students.

The comparison of different reverberation conditions and ASR (automatic subtitling systems) seems relatively novel as it appears this has not been previously compared in a teaching scenario. Two important findings were especially interesting. Firstly, reverberation has a significant impact on automatic subtitling systems, and this impacts on subjective user experience scores. Secondly, distortion destroys the ability of ASR to work effectively (detail of the results can be appreciated in Table 8).

These results are relevant because they show the importance of minimising distortion, and minimising reverberation. Ensuring that those two issues are controlled would help ensure an improvement on the staff and student experience in terms of audio. This is in-line with existing literature that emphasises the importance of audio in a learning environment on educational experience quality [22].

The results showed a correlation between subjective evaluation and ASR systems and that different types of microphones will have a different behaviour in determined conditions. To separate the systems is necessary to understand the acoustics of the room as some microphones will be more susceptible to reverberation and some to noise. The positioning and understanding of the space have been critical in the correct selection and design of the system.

The impact of reverberation on ASR and its relationship with audio quality is very relevant to the student experience, as on one hand, higher in-room/on-line audio quality directly leads to a better student experience; on the other hand, more accurate ASR will result in better accessibility to the wider student population.

A limitation of the study is that all the measurements were undertaken with an empty room. It is known from the literature than audiences will have an impact on reverberation. Although the reduction in reverberation is expected to have some small positive impact on
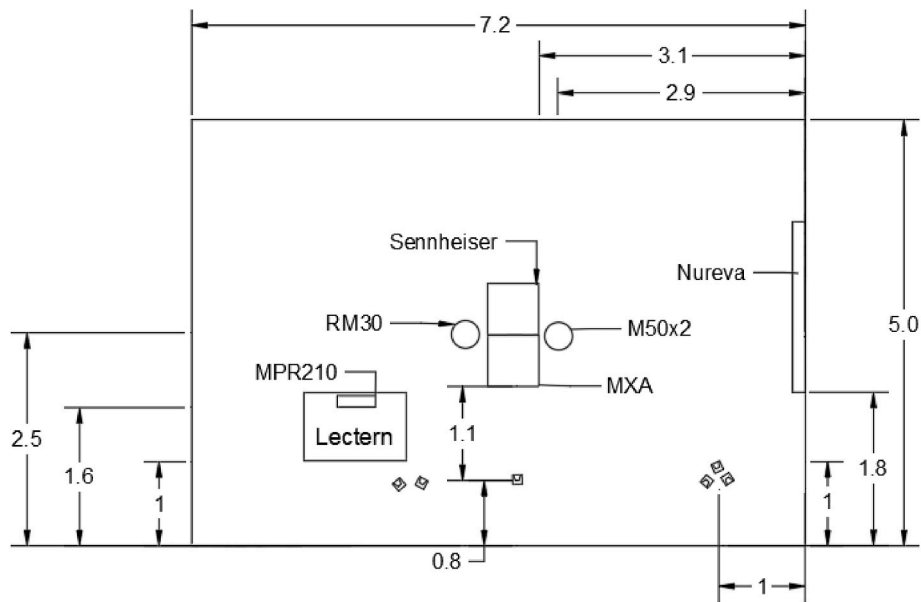
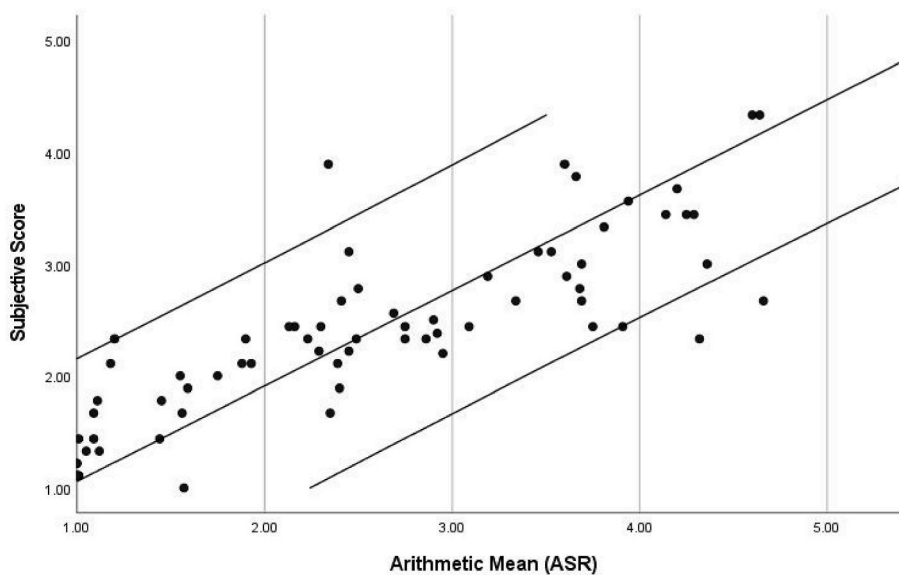**Fig. 6.** Floor plan microphones in test classroom.

Fig. 7. Standard Mean All Three ASR v Subjective Scores Scatter Plot (subjective scores presented as MOS in a 1–5 scale. ASR presented as N_{LS}).

intelligibility and clarity it is not expected that the results will vary significantly [46]. Particularly, as the rooms selected for the HyFlex implementation were pre-screened for low perceived reverberation. Similarly, the room occupancy will likely increase signal to noise ratio (SNR) variability and its impact will require further research.

The results shown are arithmetic means and as such it is important to be cognizant of the limitations of this as a technique.

It should be noted that different metrics work in different ways. Having comparable results is important; therefore employing a MOS-like scale has been used to make results comparable to Ref. [16]and as a possible extension to Ref. [17]. This also allow the results to be compared during the validation stages planned for further research.

Currently some studies focus on the quality of audio recordings [16], presented a single-ended algorithm to blindly estimate distortions from audio recordings, and VISQOL has been validated and incorporated into an open source platform [47]; however, the incorrect use of technology or understanding of the space might lead to improper implementation

and impact the student experience.

This research provides a new and simple approach to specify and integrate new technologies, such as HyFlex [3]. described in detail research on HyFlex environments but there was a gap in the research on acoustic testing. This research helps fill that gap.

Given the current research limitations and the fact that the microphones were tested in extreme reverberant conditions (5 s 0 s AKA anechoic). Care should then be taken when extrapolating these results to real life conditions and considering the implications of this work, and its effect on student satisfaction.

5.1. Further work

The influence of the use of HyFlex and their impact in students' experience as well as its variability across different disciplines and subjects are all areas of active research.

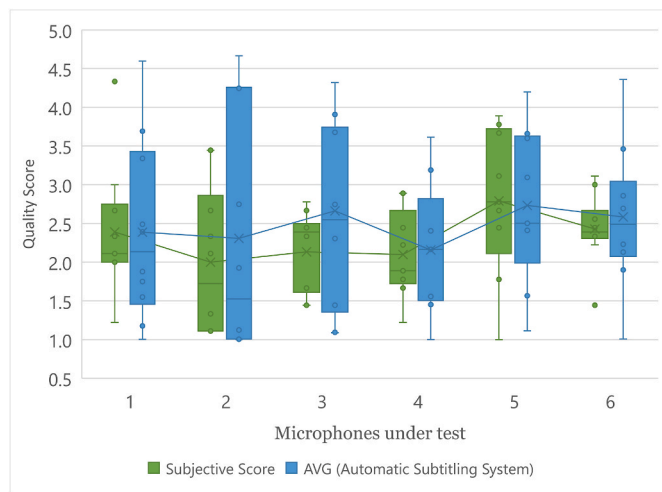The work on HyFlex is just beginning and there is scope for further

Fig. 8. ASR and subjective quality results with different microphones.

The rooms will also be analysed using an impulse response generated with a periodic pink noise and a room analyser to estimate reverberation time and other acoustic parameters.

This will allow for a comparison between recordings, the room, and the student experience. To answer the question: is there a statistically significant relation between the student experience and the audio quality as experienced in the room?

6. Conclusion

A complete design for HyFlex systems, as implemented by KCL, and general recommendations are included as part of the manuscript.

During the research 7 microphones of different types were tested in 3 different rooms with different acoustic conditions. These were evaluated and compared against each other and a reference, to catalogue their subjective sound quality. These were then used in a HyFlex setting for teaching.

The different microphone evaluations and correlations with the results from automatic subtitling seem agree with findings from Refs. [18, 19] on the impact of reverberation and ASR.

A simple methodology that uses the Levenshtein distance from ASR systems to estimate audio quality in audio installations used for teaching is presented.

- Highly reverberant spaces seem to be better serviced with beam-forming microphones with predefined lobes as opposed to those of flooded pick-up.
- Less reverberant spaces were better serviced with wide pick-up microphones instead of those with defined lobes.

validations to be undertaken in the future. The researchers are planning to evaluate the systems in place; the following is proposed:

5.1.1. Validating (ε) and implement (ζ)

After completing the installation, 5 teaching rooms will be randomly selected to test and evaluate recordings in the spaces.

The recording will be based on Harvard sentences lists from Ref. [48]. Different lists will be played back in 6 random, representative positions to fall in line with the engineering standards described by Ref. [49] with the speaker, of known directivity, positioned facing the centre of the room. The playback from the speaker is to be captured simultaneously as an uncompressed

audio file (using a portable professional digital recorder and a calibrated omni-directional microphone) and by MS Teams, which will be in use during a HyFlex session.

The recordings will be analysed using the distortion algorithm by Ref. [16] and the difference between the recorded audio will be compared using VISQOL.

Table 7
Rooms setup.

Rooms < 50 sqm	Rooms 51 − 110 sqm	Rooms > 110 sqm
1 Nureva	2 Nureva HLD 300	4-8 Shure MXA 910 & QSC DSP

Table 5
Decision matrix.

Microphone	W(sum)*0.6		W:0.3	W:0.05 Cost (1–3)	W:0.15 Portability	Total Score	Weighted
	Subjective Score	Mean (ASR)	Ease of Install (1–5)				
1	2.34	2.39	4	1	5	14.73	3.42
2	1.99	2.31	3	2	4	13.30	2.89
3	**2.54**	**2.66**	**5**	**1**	**3**	**14.21**	**3.56**
4	2.10	2.16	4	1	5	14.26	3.28
5	2.86	2.87	1	1	1	8.73	2.22
6	2.43	2.58	3	2	3	13.01	2.95
7	2.59	2.28	3	3	2	12.87	2.81
8	2.99	3.32	2	1	1	10.31	2.69

Table 6
Microphone description and mount.

	1	2	3	4	5	6	7	8
Polar Pattern	Omni Condenser	Corridor	Uniform	Omni	Multi-Lobe	Cylindrical	Cardioid	Single-Beam
Technology	Stand	Array	Mist	Cond	Beam Forming	Array	Condenser	Beam Tracking
Normal Place of installation		Desk	Wall	Ceiling	Ceiling(flushed)	Ceiling (hung)	Ceiling (inserted)	Ceiling (flushed)

- When teaching is limited to an area surrounded a fixed teaching position, microphones with a more limited pick-up area provided the best results.

This research thus provides a novel perspective on the technical implementation and on the acoustics of Hybrid flexible (HyFlex) learning systems. The results and discussion show some of the effects of different acoustic situations on microphone quality, and on commercial ASRs, all of which are likely to have an impact of student satisfaction.

Funding

A. Appendices.

All Microphones Results Box Plots

CRediT authorship contribution statement

Rodrigo Sanchez-Pizani: Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Michael Detyna:** Writing – original draft, Methodology, Investigation. **Stephen Dance:** Writing – review & editing, Validation, Supervision, Resources. **Luis Gomez-Agustina:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
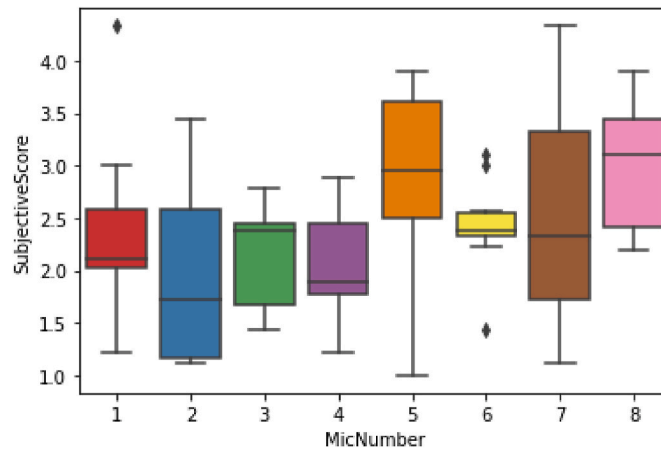

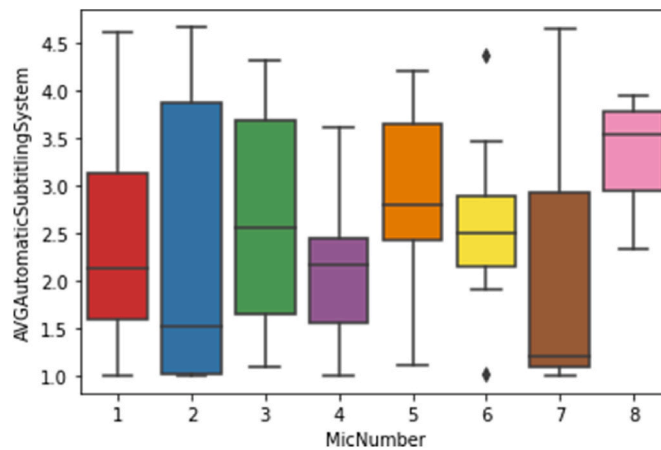**Fig. 9.** All Microphones v Subjective Scores Boxplot.

Fig. 10. All Microphones v ASR Boxplot.

All Microphones Results 1–5 rating scale

Table 8
All Microphones Results

Name	Microphone Number	Subjective Score	ASR
Earth	1	2.56	2.26
MPR	2	2.11	2.85
Nur	3	2.77	2.69
Earth2	4	1.5	1.23
MXA	5	2.78	2.57
RM30	6	2.3	2.5
ES	7	2.59	2.28
High Reverberation Only			
ES947_3 m_FR	7	1.11	1
Earth2_3 m_FR	4	1.22	1
Earth_3 m_FR	1	1.22	1
MPR_3 m_FR	2	1.33	1.12
Nureva_3 m_FR	3	1.44	1.44
RM30_3 m_FR	6	1.44	1.01
MX910_3 m_FR	5	1.78	1.11
Reverberation (H and M)			
ES947_3 m_FR	7	1.11	1
Earth2_3 m_FR	4	1.22	1
Earth_3 m_FR	1	1.22	1
MPR_3 m_FR	2	1.33	1.12
Nureva_3 m_FR	3	1.44	1.44
RM30_3 m_FR	6	1.44	1.01
Earth2_3 m_semiR	4	1.78	1.45
MX910_3 m_FR	5	1.78	1.11
Earth_3 m_SemiR	1	2.11	1.18
ES947_3 m_SemiR	7	2.33	1.2
MPR_3 m_SemiR	2	2.33	2.75
Nureva_3 m_SemiR	3	2.44	2.3
RM30_3 m_semiR	6	2.44	2.13
MX910_3 m_SemiR	5	3.11	2.45
Mean Reverberant Spaces			
Earth	1	1.67	1.09
MPR	2	1.83	1.94
Nur	3	1.94	1.87
Earth2	4	1.5	1.23
MXA	5	2.44	1.78
RM30	6	1.94	1.57
ES	7	1.72	1.1
ES	7	1.72	1.1

HyFlex Set-up, Final Tests before Deployment to rooms

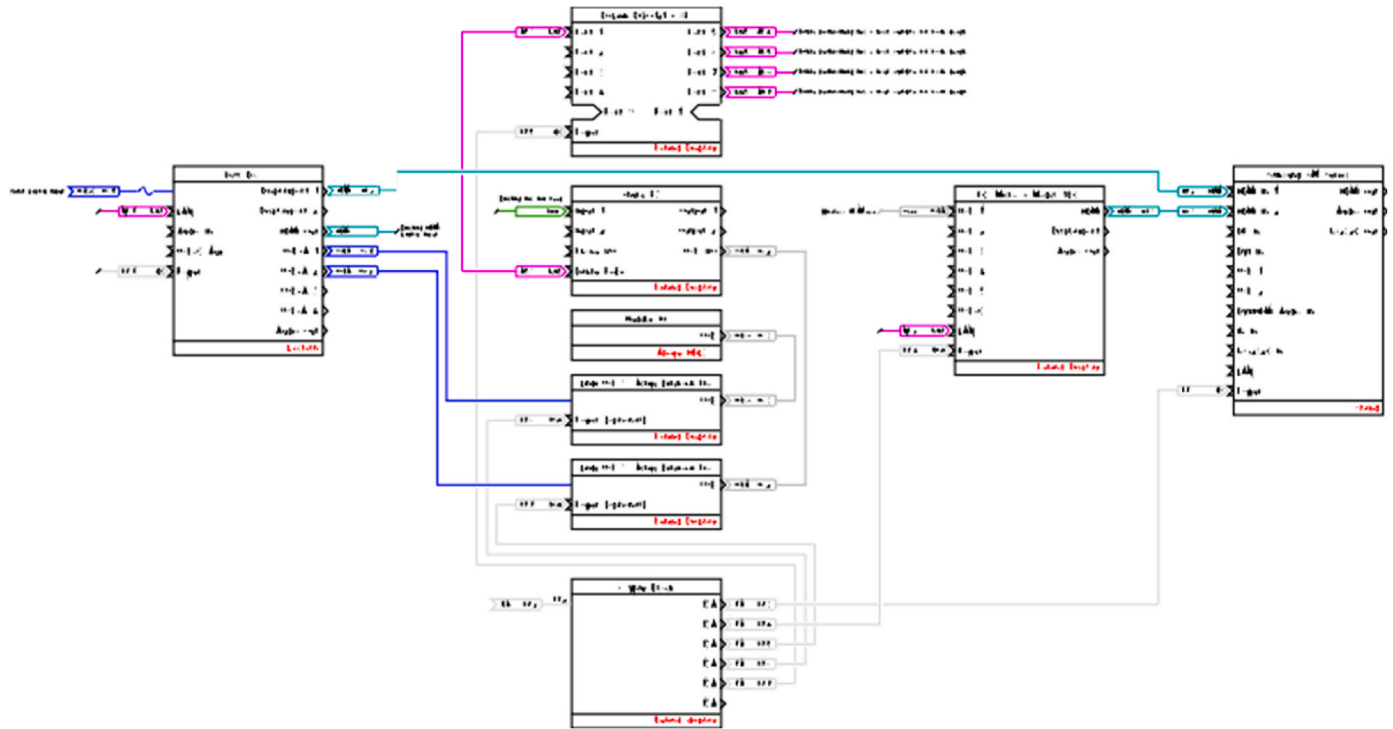**Fig. 11.** HyFlex Pilot Room Final Set-up Before Starting Tests with Volunteers.

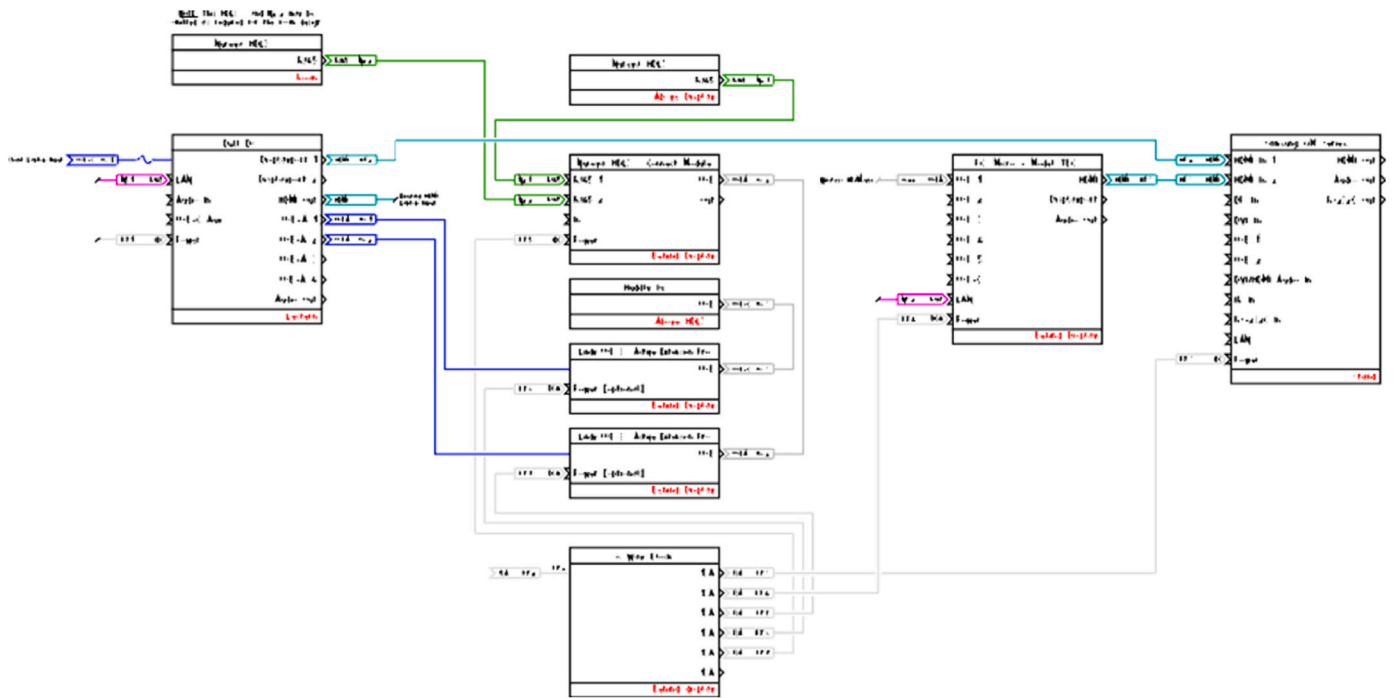**Fig. 12.** HyFlex Schematic Rooms Larger than 105m2.

Fig. 13. HyFlex Schematic Rooms Smaller than 105m2.

References

[1] March 21 Who, *Weekly Epidemiological Update on COVID-19—23 March 2021*, 2021, https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19—23-march-2021.

[2] L. Sun, Y. Tang, Z. Wei, Coronavirus pushes education online, Nat. Mater. 19 (687) (2020) 1, https://doi.org/10.1038/s41563-020-0678-8.

[3] A. Raes, L. Detienne, I. Windey, F. Depaepe, A systematic literature review on synchronous hybrid learning: gaps identified, Learn. Environ. Res. 23 (3) (2020) 269–290, https://doi.org/10.1007/s10984-019-09303-z.

[4] T. Soesmanto, S. Bonner, Dual mode delivery in an introductory statistics course: design and evaluation, J. Stat. Educ. 27 (2) (2019) 90–98, https://doi.org/10.1080/10691898.2019.1608874.

[5] D. Maher, A. Prescott, Professional development for rural and remote teachers using video conferencing, Asia Pac. J. Teach. Educ. 45 (5) (2017) 520–538, https://doi.org/10.1080/1359866X.2017.1296930.

[6] L. Cilliers, J. Pylman, South African students' perceptions of the flipped classroom: a case study of higher education, Innovat. Educ. Teach. Int. (2020) 1–10, https://doi.org/10.1080/14703297.2020.1853588, 0(0).

[7] M. Detyna, E.J. Dommett, An investigation into in-lecture digital tools for engagement: a feasibility study, Compass: J. Learn. Teach. 14 (1) (2021), https://doi.org/10.21100/compass.v14i1.1114.

[8] B.J. Beatty, Expanding the implementation of hybrid-flexible Courses and programs, in: Hybrid-Flexible Course Design, EdTech Books, 2019. https://edtechbooks.org/hyflex/adoption.

[9] M. Gosper, M. McNeill, R. Phillips, G. Preston, K. Woo, D. Green, Web-based lecture technologies and learning and teaching: a study of change in four Australian universities, ALT-J 18 (3) (2010) 251–263, https://doi.org/10.1080/09687769.2010.529111.

[10] E. Dommett, Evaluation of Lecture Capture Report April 2018. A TEL Transformation Strategy Project, KCL, 2017.

[11] N. Prodi, A. Farnetani, Y. Smyrnova, J. Fels, Investigating Classroom Acoustics by Means of Advanced Reproduction Techniques, vol. 10, 2007.

[12] A.-M. Lafortune, S. Lakhal, Differences in students' perceptions of the community of inquiry in a blended synchronous delivery mode, Can. J. Learn. Technol./La Revue Canadienne de l'apprentissage et de La Technologie 45 (3) (2020), https://doi.org/10.21432/cjlt27839.

[13] M. Lavandier, P. Herzog, S. Meunier, Comparative measurements of loudspeakers in a listening situation, J. Acoust. Soc. Am. 123 (1) (2008) 77–87, https://doi.org/10.1121/1.2816571.

[14] J.G. Beerends, M. Obermann, R. Ullmann, J. Pomy, M. Keyhl, Perceptual objective listening quality assessment (POLQA), the third generation ITU-T standard for end-to-end speech quality measurement Part I–temporal alignment, J. Audio Eng. Soc. 61 (6) (2013) 19.

[15] T. ITU, *P.863 Series P: Telephone Transmission Quality, Telephone Installations, Local Line Networks. Perceptual Objective Listening Quality Prediction* (P:Telephone Transmission Quality, Telephone Installations, Local Line Networks.), International Telecommunication Union, 2018.

[16] P. Kendrick, I. Jackson, F.F. Li, T.J. Cox, B. Fazenda, Perceived audio quality of sounds degraded by nonlinear distortions and single-ended assessment using HASQI, J. Audio Eng. Soc. 63 (9) (2015) 698–712, https://doi.org/10.17743/jaes.2015.0068.

[17] M. Schoeffler, S. Bartoschek, F.-R. Stöter, M. Roess, S. Westphal, B. Edler, J. Herre, WebMUSHRA — a comprehensive framework for web-based listening tests, J. Open Res. Software 6 (2018) 8, https://doi.org/10.5334/jors.187.

[18] A. Brutti, M. Matassoni, On the relationship between early-to-late ratio of room impulse responses and ASR performance in reverberant environments, Speech Commun. 76 (2016) 170–185, https://doi.org/10.1016/j.specom.2015.09.004.

[19] M. Wolf, C. Nadeu, Channel selection measures for multi-microphone speech recognition, Speech Commun. 57 (2014) 170–180, https://doi.org/10.1016/j.specom.2013.09.015.

[20] J.M. Zydney, P. McKimmy, R. Lindberg, M. Schmidt, Here or there instruction: lessons learned in implementing innovative approaches to blended synchronous learning, TechTrends 63 (2) (2019) 123–132, https://doi.org/10.1007/s11528-018-0344-z.

[21] L. Scannell, M. Hodgson, J. García Moreno Villarreal, R. Gifford, The role of acoustics in the perceived suitability of, and well-being in, informal learning spaces, Environ. Behav. 48 (6) (2016) 769–795, https://doi.org/10.1177/0013916514567127.

[22] Z. Yang, B. Becerik-Gerber, L. Mino, A study on student perceptions of higher education classrooms: impact of classroom attributes on student satisfaction and performance, Build. Environ. 70 (2013) 171–188, https://doi.org/10.1016/j.buildenv.2013.08.030.

[23] F. Pausch, Z.E. Peng, L. Aspöck, J. Fels, Speech perception by children in a real-time virtual acoustic environment with simulated hearing aids and room acoustics, Free-Field.Virtual.Psychoacoustics. Hear.Impair. 10 (2016).

[24] H.J.M. Steeneken, T. Houtgast, A physical method for measuring speech-transmission quality, J. Acoust. Soc. Am. 67 (1) (1980) 318–326, https://doi.org/10.1121/1.384464.

[25] L. Morales, et al., Speech transmission index for the English language verified under reverberant conditions with two binaural listening methods: real-life and headphones, J. Audio Eng. Soc. 62 (7/8) (2014) 493–504, https://doi.org/10.17743/jaes.2014.0029.

[26] L. Morales, G. Leembruggen, S. Dance, B.M. Shield, A revised speech spectrum for STI calculations, Appl. Acoust. 132 (2018) 33–42, https://doi.org/10.1016/j.apacoust.2017.11.008.

[27] BSI, BS EN IEC 60268-16-2020. Sound System Equipment. Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index, 2020. BSI Standards Limited 2020.

[28] Great Britain & School Buildings and Design Unit, Acoustic Design of Schools: A Design Guides, TSA, 2003.

[29] R. Daniels, *Acoustic Design of Schools: Performance Standards. BB93*. Departmet for Education, 2015. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/400784/BB93_February_2015.pdf.

[30] Y. Shang, Expert systems, in: The Electrical Engineering Handbook, Elsevier, 2005, pp. 367–377, https://doi.org/10.1016/B978-012170960-0/50031-1.

[31] M. Guenther, Blurring boundaries, in: Intersection, Elsevier, 2013, pp. 36–61, https://doi.org/10.1016/B978-0-12-388435-0.50002-0.

[32] J. Beerends, I. Beerends, On the assessment of high-quality voice recordings including voice postprocessing, J. Audio Eng. Soc. 63 (3) (2015) 174–183, https://doi.org/10.17743/jaes.2015.0013.

[33] Z. Su, B. Ahn, K. Eom, M. Kang, J. Kim, M. Kim, Plagiarism detection using the Levenshtein distance and smith-waterman algorithm, in: 2008 3rd International Conference on Innovative Computing Information and Control, 2008, https://doi.org/10.1109/ICICIC.2008.422, 569–569.

[34] M. Koskinen, Understanding the Needs of Adult Graduate Students: an Exploratory Case Study of A Hyflex Learning Environment, [Doctor of Education, Northeastern University, 2018. https://repository.library.northeastern.edu/files/neu:m044c738v/fulltext.pdf.

[35] November 23 L. Gogia, HyFlex ByeFlex: Rethinking the Way Forward, The Tambellini Group, 2020, https://www.thetambellinigroup.com/hyflex-byeflex-rethinking-the-way-forward/.

[36] G. Jancke, J. Grudin, A. Gupta, Presenting to Local and Remote Audiences: Design and Use of the TELEP System, vol. 10, Microsoft Research, 1999.

[37] BSI, Sound System Equipment—Part 4: Microphones (IEC 60268-4:2014) Equipements, British Standards Institution, 2014.

[38] February T. ITU, *ITU-T Test Signals For Telecommunication Systems.Test Vectors Associated To Rec. ITU-T P.50 Appendix I*. ITU, 1998, http://www.itu.int/net/itu-t/sigdb/genaudio/AudioForm-g.aspx?val=1000050.

[39] J.S. Bradley, B.N. Gover, Subjective and Objective Rating of Intelligibility of Speech Recordings, 2007, p. 12. *Convention Paper 7178*.

[40] B.M. Fazenda, P. Kendrick, T.J. Cox, F. Li, I. Jackson, Perception and Automated Assessment of Audio Quality in User Generated Content, vol. 5, 2015. New York.

[41]] Workers, N. R. C. (US) P. on I. of V. V. on V. of, & Vision, N. R. C. (US) C. on, Lighting and reflections, in: Video Displays, Work, and Vision, National Academies Press (US), 1983. https://www.ncbi.nlm.nih.gov/books/NBK216496/.

[42] AVIXA, Display Image Size for 2D Content in Audiovisual Systems, 2016. ANSI/INFOCOMM V202.01:2016. InfoCommInternational.

[43] ISO, BSI, B, BS ISO 517:2008.Photography—Apertures and Related Properties Pertaining to Photographic Lenses—Designations and Measurements, British Standard Institution, 2008. *ISO 517*.

[44] T. ITU, SERIES F: NON-TELEPHONE TELECOMMUNICATION SERVICES, Multimedia Services, Service Requirements for Next Generation Content Delivery Networks, 2018. F.743.6. International Telecommunication Union.

[45] D. Sotirios, C. George, The acoustics of learning environments and implications in communication and learning, The 2005 Congress and Exposition on Noise Controlntrol 11 (2005).

[46] Y.-J. Choi, Effect of occupancy on acoustical conditions in university classrooms, Appl. Acoust. 114 (2016) 36–43, https://doi.org/10.1016/j.apacoust.2016.07.010.

[47] M. Chinen, F.S.C. Lim, J. Skoglund, N. Gureev, F. O'Gorman, A. Hines, ViSQOL V3: an Open Source Production Ready Objective Speech and Audio Metric, 2020. *ArXiv:2004.09584 [Cs, Eess]*, http://arxiv.org/abs/2004.09584.

[48] March 1 C. Hopkins, S. Graetzer, G. Seiffert, ARU Speech Corpus (University of Liverpool), University of Liverpool, 2019 [Data Collection], http://datacat.liverpool.ac.uk/681.

[49] BSI, BS EN ISO 3382-1:2009, British Standards Institution, 2009.

[50] ITU, T. (02/96).Methods for Objective and Subjective Assesment of Quality. Subjective Performance Assesment of Telephone-Band and Wideband Digital Codecs. ITU P.830. International Telecommunication Union.