Student activity and speech levels before and after acoustic enhancement and PA redesign

Domenico De Salvio[1*], **Dario D'Orazio**[1], **Massimo Garai**[1]

[1]Department of Industrial Engineering, University of Bologna
Viale del Risogimento 2, 40136, Bologna, Italy

*{domenico.desalvio2@unibo.it}

Abstract
It is largely known that student activity (SA) - i.e. the noise generated by students during lessons - deeply affects the learning performance. Moreover, teachers' speech level (SL) may depend on vocal support, the latter being due in large rooms both to acoustic properties of the room and PA system. Previous work showed that clustering techniques can be a useful tool to measure SA and SL through long-term monitoring of active classrooms. In the present work, SA and SL measured in two historical university lecture halls are compared before and after renovation works. Restoration included both acoustic treatments and PA re-design. It is worth noting that measurements were carried out in a pre-COVID19 scenario. Two unsupervised machine learning algorithms were used, respectively Gaussian Mixture Model and K-means clustering. Outcomes have been compared with equivalent and percentile levels, usually used in this field of analysis. Results show lower mean levels of both SA and SL and lower signal-to-noise ratios, suggesting the achieving of quieter environments. Furthermore, the decrease of the Lombard slope hints at better control of the vocal effort by the teachers.

Keywords: speech intelligibility, student activity, speech level, Gaussian Mixture model, K-means clustering.

1 Introduction

The quality of the learning process is strictly dependent on intelligibility. The latter is outlined as a function of the acoustic characteristics of the room and the amount of background noise [1]. The first factors can be precisely measured by objective metrics, such as the reverberation time, the early-to-late index, and the Speech Transmission Index.

However, the background noise is not well-defined when the dynamical context of a lecture is considered. Noise during lessons comes from human activity, either in the neighbouring areas of the lecture halls or by the students attending lectures. The latter is called student activity [2]. Both factors, acoustic criteria and background noise within the space, can give rise to the Lombard effect, the psychoacoustic mechanism that leads the speakers to increase their voice levels [3]. Thus, it brings higher vocal efforts [4].

As a consequence of what outlined so far, occupancy and public address can play a key role in this dynamic context. The first changes the equivalent absorption area as a function of the number of students and their placement [5]; the second allows the speakers to reach high signal-to-noise ratios without strain the voice [6].

Active classrooms can be measured, besides the classical percentile and continuous equivalent levels, through the post-processing of the recording of a sound level meter via clustering techniques, such as Gaussian Mixture Model or K-means clustering [2, 7, 8].

The present work compares active classrooms measurements carried out in two historical lecture halls before and after acoustic treatments. The aim is to investigate a possible change in behaviour of both students and teachers in treated spaces and the chance to use these methods as a tool to assess the effects of the renovation works.

2 Method

The assessment of renovation works carried out in two historical lecture halls (see Fig. 1) has been made through monitoring of the student activity and speech levels before and after the treatments. The two lecture halls are very similar, with a rectangular plan and seats made of wood with an amphitheater shape. The main difference concerns an extra volume of about 100 m^3 in the rear part of Hall I. As a consequence, Hall I contains 250 students, whereas Hall II has 200.

(a) Lecture hall I   (b) Lecture hall II

Figure 1: Inner view of the two lecture halls under study before the treatments.

Renovation works concerned two interventions: the acoustic treatment of the surfaces, and the redesign of the PA. Regarding the first, slat absorbers wooden panels have been placed in the rear walls and on the overhanging beams of both rooms. The PA has been replaced with two systems: a main line array on the back of the teacher on either side of the blackboard, and a pair of fillers in the middle of the room to broader the coverage homogeneously over the audience area.

Objective parameters, according to ISO 3382, were measured in both lecture halls in unoccupied conditions using the same source-receiver pairs for both before and after treatments state. The comparison focuses on the reverberation time T_{30}, the early-to-late index C_{50}, and the Speech Transmission Index STI. Table 1 shows the impact of the treatments besides the general and geometrical characteristics of the halls.

Table 1: General and acoustic data of the halls under study before and after the restoration, respectively indicated as "ante" and "post". Besides the shape of the inner space, it is shown the volume "V", the maximum occupancy "N", the reverberation time in unoccupied state "T", the early-to-late index "C_{50}", the Speech Transmission Index STI and the equivalent absorption area A_0 of the lecture halls in unoccupied state. The subscript "M" states a value averaged over all the receivers in the octave bands of 500 – 1000 Hz, whereas "3" over the octave bands of 500 – 2000 Hz.

Hall	Shape	Volume (m^3)	Occupancy	T_M (s)		$C_{50,3}$ (dB)		STI		A_0 (m^2)	
		V	N	Ante	Post	Ante	Post	Ante	Post	Ante	Post
I	Amphitheater	1000	250	1.70	1.37	-2.8	-1.4	0.49	0.52	94	117
II	Amphitheater	900	200	1.72	1.38	-2.4	-1.0	0.47	0.54	84	105

2.1. Student activity and speech levels method

Basing on previous works, student activity SA and speech levels SL were measured through two sound level meters placed on either side in the middle of the audience area, far enough from reflective surfaces. Monitoring regarded 9 lessons before and after the renovation works. The method previously described in [7] was used.

The sound level meters recorded equivalent continuous levels with an acquisition time of 0.1 s to ensure the recording of the pauses within the speech's streaming [9].

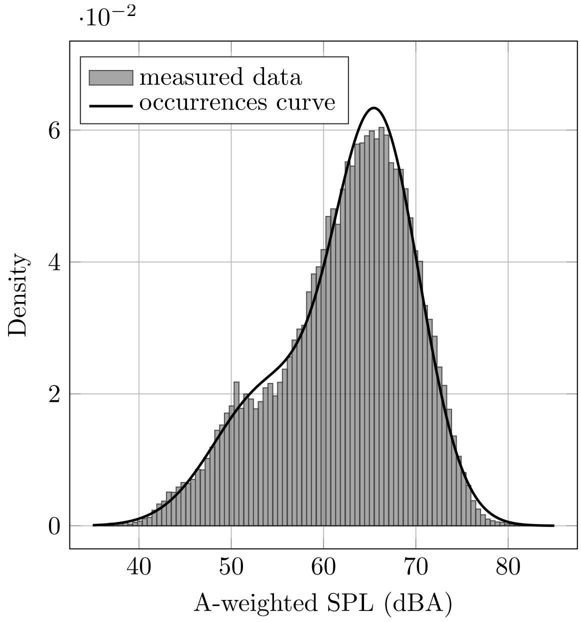


Figure 2: Example of the occurrences curve obtained by a recorded lesson. The histogram shows the normalized bin count of the A-weighted recorded SPLs, whereas the curve indicates the probability distribution function to process via clustering techniques.

In post-processing, time histories were cut to analyze and process only the lecture time. Then, two unsupervised algorithms, i.e. Gaussian Mixture Model GMM and K-means clustering KM, found patterns among the recorded SPLs and gathered them into different clusters.

Briefly, the GMM is a model-based clustering algorithm that splits the occurrences distribution of the SPLs as a sum of Gaussian curves according to statistical hypotheses. Each cluster - thus, each sound source - is represented by a Gaussian curve, and the corresponding mean is assumed as the SPL of the sound source. KM is a distance-based clustering algorithm that splits the recorded SPLs minimizing and optimizing the distance among data. Each cluster - associated with each sound source, as seen for GM - is represented by a bunch of SPLs with a corresponding "centroid", i.e. the centre of gravity of the cluster, which is assumed as the SPL of the sound source.

Being the acoustic environment compounded by two sound sources, the number of clusters to run the algorithms was set equal to 2. Then, the lower means or centroids of clusters were assigned to SA and the higher values to SL.

3 Results and discussions

The analysis concerns a total of 18 lessons, 9 recorded before the acoustic treatments of the halls and 9 after. Table 2 shows the outcomes of the clustering process. Besides the SPLs obtained for SA and SL, other data are the corresponding hall, the occupancy, the equivalent absorption area in the occupied state, and the 90th percentile and the continuous equivalent level. Brackets contain the standard deviations between the two sound level meters.

Table 2: Overview of the recorded lessons. For each lesson, the corresponding room, the number of people, the percentage of occupancy, and the equivalent absorption area taking into account the contribution of the people are shown. Measured A-weighted values of student activity SA, received speech level SL extracted through, Gaussian Mixture Model GMM, K-means clustering KM, equivalent continuous L_{eq} and percentile levels L_{90} are reported. Values are averaged over the two receiver positions. All values of SA and SL are in dBA.

Lesson	Hall	Occupancy (%)	A_{occ} (m^2)	GMM		KM		L_{90}	L_{eq}
				SA (s.d.)	SL (s.d.)	SA (s.d.)	SL (s.d.)	(s.d.)	(s.d.)
A	I (Ante)	145 (60%)	156	48.2 (1.2)	65.0 (4.0)	52.2 (1.2)	68.3 (4.0)	48.0 (0.5)	69.9 (4.8)
B	I (Ante)	200 (80%)	179	47.5 (1.5)	63.3 (4.6)	48.8 (1.3)	64.2 (4.5)	45.8 (0.6)	64.8 (4.9)
C	I (Ante)	100 (50%)	137	53.3 (1.8)	66.3 (4.1)	55.8 (1.9)	68.4 (3.9)	53.0 (1.7)	68.9 (4.5)
D	I (Ante)	150 (60%)	158	51.2 (2.0)	67.2 (4.5)	52.7 (1.9)	68.4 (4.4)	47.6 (1.3)	69.7 (4.6)
E	II (Ante)	250 (125%)	190	48.4 (0.3)	67.5 (1.5)	49.1 (0.1)	68.0 (1.6)	52.1 (9.5)	72.3 (7.1)
F	II (Ante)	160 (80%)	152	50.3 (1.5)	66.5 (1.5)	53.1 (0.2)	68.5 (1.0)	55.0 (8.1)	71.9 (4.0)
G	II (Ante)	120 (60%)	135	61.0 (0.6)	75.5 (1.4)	55.7 (0.7)	74.9 (1.4)	61.6 (5.4)	78.6 (5.3)
H	II (Ante)	150 (75%)	164	55.3 (0.1)	75.3 (0.8)	55.8 (0.0)	76.0 (0.8)	56.4 (7.0)	79.2 (3.6)
I	II (Ante)	200 (100%)	188	53.4 (0.0)	68.0 (1.0)	56.5 (0.3)	69.7 (0.8)	58.8 (5.9)	74.6 (3.7)
Mean	Ante	164	162	52.1 (1.0)	68.3 (2.6)	53.3 (0.8)	69.6 (2.5)	53.1 (4.4)	72.2 (4.7)
J	I (Post)	130 (50%)	172	51.3 (1.1)	64.6 (0.7)	52.7 (0.7)	66.1 (0.5)	49.7 (0.8)	66.5 (0.4)
K	I (Post)	185 (75%)	195	49.9 (0.4)	69.6 (0.4)	50.9 (0.4)	70.2 (0.3)	48.1 (0.5)	71.5 (0.1)
L	I (Post)	130 (50%)	172	47.2 (0.5)	64.6 (0.5)	50.9 (0.6)	70.2 (0.5)	45.9 (0.4)	65.8 (0.4)
M	I (Post)	80 (30%)	151	49.9 (0.4)	59.0 (0.3)	51.0 (0.4)	61.2 (0.2)	48.3 (0.5)	61.1 (0.0)
N	I (Post)	190 (75%)	197	52.7 (0.4)	68.9 (0.1)	51.0 (0.2)	69.0 (0.1)	48.8 (0.5)	70.3 (0.1)
O	II (Post)	110 (55%)	168	53.9 (3.1)	72.1 (1.1)	54.1 (1.3)	72.7 (2.0)	53.3 (1.0)	74.4 (2.4)
P	II (Post)	125 (65%)	175	51.4 (0.8)	60.8 (1.3)	50.4 (0.9)	62.0 (2.1)	46.7 (0.7)	62.0 (1.3)
Q	II (Post)	120 (60%)	173	48.8 (0.3)	66.4 (1.1)	49.7 (0.9)	69.8 (2.1)	58.4 (0.5)	71.6 (3.0)
R	II (Post)	95 (50%)	161	52.1 (1.8)	64.9 (1.9)	54.4 (1.5)	67.0 (2.1)	51.4 (1.1)	67.5 (2.3)
Mean	Post	129	174	50.8 (1.0)	65.6 (0.8)	51.7 (0.8)	67.6 (1.1)	50.0 (0.7)	67.8 (1.1)

The measured A-weighted SA and SL values before the treatments lie respectively in the range of 47.5 – 61 dB and 63.3 – 75.5 dB for GMM, 48.8 – 56.5 dB and 64.2 – 76 dB for KM, 45.8 – 61.6 dB and 64.8 – 79.2 for percentile and equivalent levels. The measured SA and SL levels after the treatments lie respectively in the ranges 47.2 – 53.9 dB and 59 – 72.1 dB for GMM, 49.7 – 54.1 dB and 61.2 – 72.7 dB for KM, 45.9 – 53.3 dB and 61.1 – 74.4 dB for percentile and equivalent levels. Before the restoration work, the standard deviations between the two receivers, respectively for SA and SL, lie in the ranges 0 – 2 and 0.8 – 4.6 dB for GMM, 0 – 1.9 and 0.8 – 4.5 dB for KM, 0.5 – 9.5 and 3.6 – 7.1 dB for percentile and equivalent levels. Concerning the measured s.d. of SA and SL after the treatments, values lie respectively in the ranges 0.3 – 3.1 and 0.1 – 1.9 dB for GMM, 0.2 – 1.5 and 0.1 – 2.1 dB for KM, 0.4 – 1.1 and 0 – 3 dB for percentile and equivalent levels. The measured A-weighted mean values of SA and SL and their standard deviations in brackets before the treatments are respectively 52.1 (1.0) and 68.3 (2.6) dB for GMM, 53.3 (0.8) and 69.6 (2.5) dB for KM, 53.1 (4.4) and 72.2 (4.7) dB for percentile and equivalent levels. The same parameters measured after the treatments are respectively 50.8 (1.0) and 65.6 (0.8) dB for GMM, 51.7 (0.8) and 67.6 (1.1) dB for KM, 50 (0.7) and 67.8 (1.1) for percentile and equivalent levels.

3.1. SA, SL and algorithms

The first analysis concerns the comparison of SA and SL obtained before (lessons A - I) and after (lessons J - R) the treatments. Means and standard deviations are lower for all techniques and all parameters; thus, it could mean that the lecture halls have quieter and more diffused sound environments. The greatest decrease of both metrics, i.e. means and s.d., concerns the percentile and equivalent levels.

According to the results of previous work [7], KM has the higher values of SA in the greater part of the lectures and the 90th percentile the lowest. Concerning SL, equivalent levels are the highest measured values and the GMM the lowest in all lectures.

Other issues concerning GMM and KM are about the heteroscedasticity and the initial hypotheses of the two techniques. Heteroscedasticity is the difference of variance among data. The greater the variance, the greater the difference between GMM and KM [10]. The initial hypotheses regard the difference between hard and soft clustering. GMM is a soft clustering algorithm since it allows the data to belong to one or more clusters with an assigned probability. By contrast, KM is a hard clustering algorithm and each data point is assigned only to one cluster [11]. Other applications of these methods pointed out the difference concerning the initial hypotheses between these two algorithms [12].

3.2. Signal-to-noise ratio and Lombard effect

In the present work, the signal-to-noise ratio is defined as the difference between SL, the signal, and SA, the noise. As seen in the previous section and reported in Table 2, SNRs are slightly lower of an average of about 1 dB after the treatments. Figure 3 plots the relationship between SA and SL before (black marks and lines) and after (red marks and lines) the treatments. Here, the slopes of the regressions show changes in the Lombard effect. That is because the decrease of SA and SL is not proportional. The decrease of the relationship between SA and SL could mean that the vocal effort of teachers is less affected by the babble of the students and the Lombard effect is differently triggered.

3.3. The role of the occupancy

In university lecture halls, the occupancy is not constant during the day but changes every lesson. Thus, the acoustic characteristics of the halls change dynamically basing on the students attending lessons. In the prediction model built up by Hodgson, the equivalent absorption area of the room in occupied condition is one of the main parameters [2]. Figure 4 shows the relationship between occupancy and SNR, SA, and SL. Black and red lines represent respectively the relationship before and after the restoration in all plots. In the left part, the relationship between occupancy and the SNR is shown. The enhancement of the acoustic conditions of the halls seems to make the correlation more sensitive. SNR increases linearly with the occupancy, indeed. This could mean that the bigger the audience, the quieter the environment.

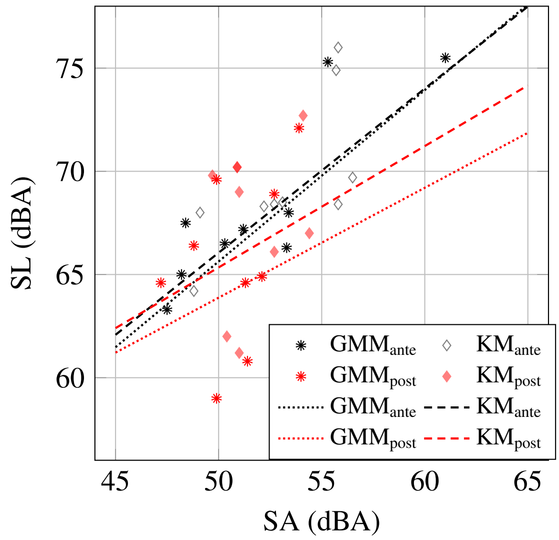


Figure 3: Relationship between SA and SL measured values via GMM and KM. Black and red markers indicate respectively before and after the acoustic treatments. Each marker refers to a whole lesson.

An occupancy of about 120 students seems to be a kind of threshold according to the relationship with the SNR. When audiences are greater than 120 people seem to keep an SNR equal or higher of 15 dB, whereas lectures attended by less than 120 students were carried out with an SNR lower than 15 dB, down to 9 dB. Considering the size of large lecture halls, a PA seems to be necessary to achieve good intelligibility without affecting the vocal effort of teachers.

Being the SNR depending on SA and SL, it is worth analyzing how these two factors are related to the occupancy. Thus the middle and the right plots in Figure 4 seem to suggest how treatments change the behaviour of both students and teachers. SA seems to be independent of the occupancy after the renovation works. Regression lines are almost constant for both algorithms, indeed. Concerning SL, regression lines change slopes, making the vocal intensity of teachers more sensitive to the occupancy after the treatments.

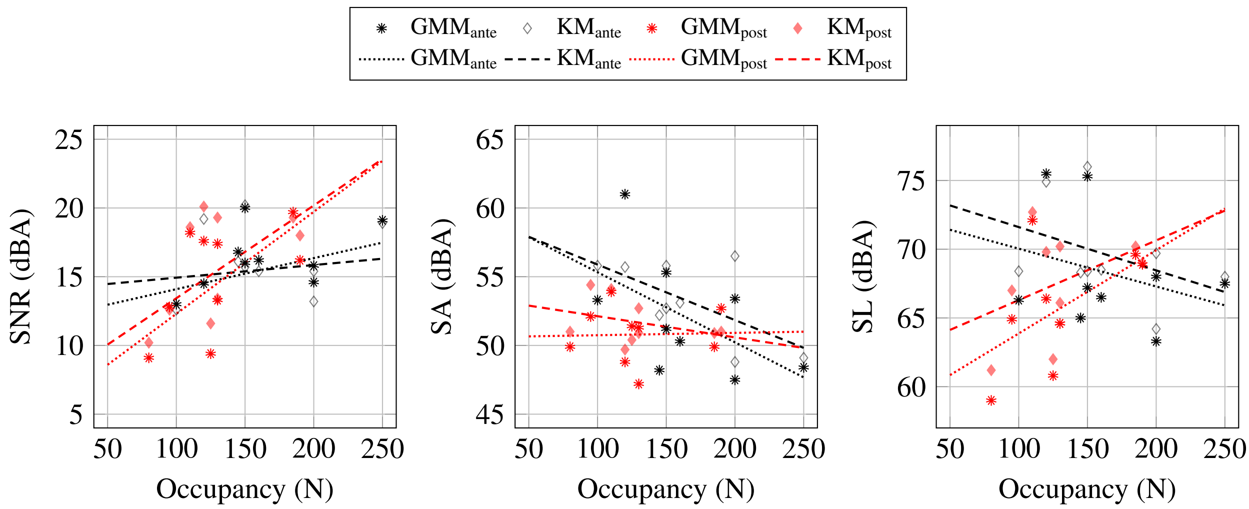
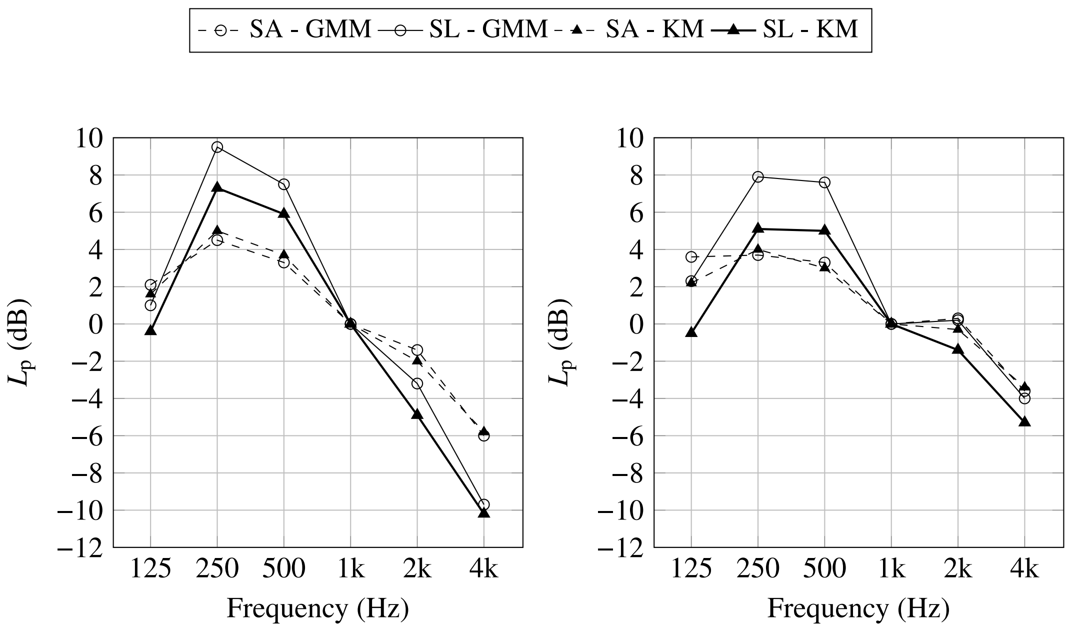


Figure 4: Relationship between Occupancy and SNR, SA and SL measured values via GMM and KM. Black and red markers indicate respectively before and after acoustic treatments. Each marker refers to a whole lesson.

3.4. Spectral analysis

The investigation of the spectra obtained via GMM and KM may provide further evidence about the reliability of clustering techniques. Being both sound sources are developed by speech signals, it is expected to obtain two speech shapes. Spectral analyses were made applying GMM and KM on the recorded SPLs in the octave band from 125 to 4000 Hz. Figure 5 shows the results for both before and after the treatments, respectively in Figures 5a and 5b. Relative spectra are plotted setting the 1 kHz octave band equal to 0 averaging over all measured lectures.

The shapes obtained are attributable to speech sources. This is quite clear for SL before treatments, whereas SA has flatter shapes being more diffuse through space. Different results concern the state after treatments where similar spectra were obtained from low frequencies up to 1 kHz octave band. The highest frequencies show weakness in separating the two sound sources. The 2 and 4 kHz octave bands seem to be influenced by the new PA which emphasizes the SL mid-high frequencies. SA can not be affected by the properties of the PA. Thus, clustering seems to be less reliable after the treatments. Reasons can be assumed on multiple sides. Concerning the algorithms, it should be noted that the more visible are the peaks of the occurrences curve, the easier is the detection of different clusters. Moreover, treatments regarded especially mid-high frequencies, hence sound energy of formants' speech is weaker within the room and less detectable. Finally, according to the outcomes shown in previous sections, the hypothesis of achieving quieter environments seems to be confirmed.



(a) Ante - relative spectra of SA and SL

(b) Post - relative spectra of SA and SL

Figure 5: Average relative spectra of student activity SA and speech levels SL obtained via GMM and KM. On the left SA and SL obtained before the acoustic treatments of the halls are shown, on the right the after state. Values are averaged over all measured lectures.

3.5. Monitoring SA and SL during lessons

Data population of each lecture were divided in 15 minutes samples to investigate the temporal fluctuations of SA and SL during lessons. Figure 6 shows the measured tendencies for both before and after treatments and both algorithms. Differences in this kind of analysis are strictly related to the shape of the occurrences curve and the consequent variance of data. Results are quite coherent between the two algorithms, however KM seems to be more robust and less affected by fluctuations. The increase or decrease of SA during time could concern the listening effort of the students and the consequent concentration. The up and down tendency is particular evident in lesson H recorded before the renovation works.

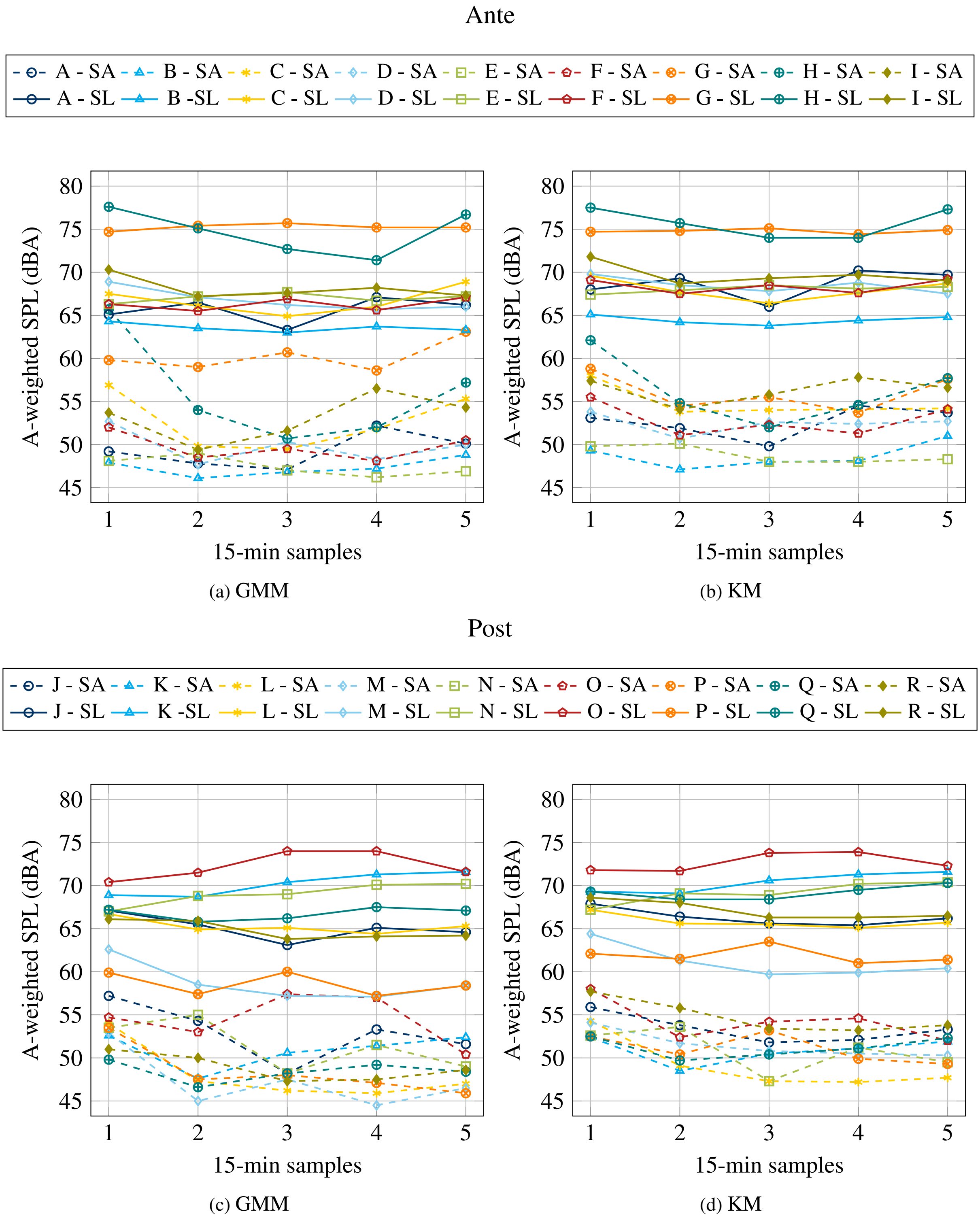


Figure 6: 15-minutes samples of SA and SL for each lecture before and after the acoustic treatments of the halls. SA and SL are indicated respectively with dashed and solid lines.

4 Conclusions

Good acoustic conditions of environments affect the behaviour of teachers and students within lecture halls. Measurements of student activity and speech levels were carried out during 18 lessons to test the effectiveness of acoustic treatments and audio redesign of two university lecture halls. Eighteen lessons were recorded through two sound level meters and analyzed via clustering techniques: Gaussian Mixture Model and K-means. Results show decreases in mean values of student activity and teachers' speech levels besides standard deviations between the two sound level meters. Thus, it is conceivable that treatments led to quieter and more diffused sound environments. Drops of Lombard effect slopes show how speech levels are less affected by student activity. Treatments influenced the behaviour of students and teachers investigating the relationships among occupancy, signal-to-noise ratio, student activity, and speech levels. After the renovation works, speech levels seem to be more sensitive to occupancy, whereas student activity is less dependent on the number of people attending lectures. The spectral analysis confirmed the reliability of clustering techniques. However, it showed weaknesses when the peaks of the occurrences curves are not clear. A further analysis concerned the chance to use clustering algorithms to monitor student activity and speech levels during lectures. Thus, 15-minutes samples were analyzed via Gaussian Mixture Model and K-means to assess the vocal effort of teachers and the concentration of students. Results show how student activity and speech levels fluctuate during lecture time, highlighting possible listening efforts by students. The assessment of the effectiveness of renovation works made through clustering techniques confirms the reliability of these algorithms. Further works should aim to dig into quantitative analyses and make the use of these unsupervised approaches more robust.

References

[1] Murray Hodgson. Experimental investigation of the acoustical characteristics of university classrooms. *The Journal of the Acoustical Society of America*, 106(4):1810–1819, 1999.

[2] M Hodgson, R Rempel, and S Kennedy. Measurement and prediction of typical speech and background noise levels in univiersity classrooms during lectures. *The Journal of the Acoustical Society of America*, 105(1):226–233, 1999.

[3] Etienne Lombard. Le signe de l'elevation de la voix. *Ann. Mal. de L'Oreille et du Larynx*, pages 101–119, 1911.

[4] Jonas Brunskog, Anders Christian Gade, Gaspar Payá Bellester, and Lilian Reig Calbo. Increase in voice level and speaker comfort in lecture rooms. *The Journal of the Acoustical Society of America*, 125(4):2072–2082, 2009.

[5] Young-Ji Choi. Effect of occupancy on acoustical conditions in university classrooms. *Applied Acoustics*, 114:36–43, 2016.

[6] Jianxin Peng, Honghu Zhang, and Dan Wang. Measurement and analysis of teaching and background noise level in classrooms of chinese elementary schools. *Applied Acoustics*, 131:1–4, 2018.

[7] Dario D'Orazio, Domenico De Salvio, Laura Anderlucci, and Massimo Garai. Measuring the speech level and the student activity in lecture halls: Visual-vs blind-segmentation methods. *Applied Acoustics*, 169:107448, 2020.

[8] Lily M Wang and Laura C Brill. Speech and noise levels measured in occupied k–12 classrooms. *The Journal of the Acoustical Society of America*, 150(2):864–877, 2021.

[9] Dagen Wang and Shrikanth S Narayanan. Robust speech rate estimation for spontaneous speech. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(8):2190–2201, 2007.

[10] David Barber. *Bayesian reasoning and machine learning*. Cambridge University Press, 2012.

[11] Michael J Bianco, Peter Gerstoft, James Traer, Emma Ozanich, Marie A Roch, Sharon Gannot, and Charles-Alban Deledalle. Machine learning in acoustics: Theory and applications. *The Journal of the Acoustical Society of America*, 146(5):3590–3628, 2019.

[12] Domenico De Salvio, Dario D'Orazio, and Massimo Garai. Unsupervised analysis of background noise sources in active offices. *The Journal of the Acoustical Society of America*, 149(6):4049–4060, 2021.