PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Signal-to-noise ratio in university lecture halls with low intelligibility

Dario D'ORAZIO(1), Domenico DE SALVIO(2), Laura ANDERLUCCI(3), Massimo GARAI(4)

(1)DIN, Department of Industrial Engineering, University of Bologna, Italy, dario.dorazio@unibo.it

(2)DIN, Department of Industrial Engineering, University of Bologna, Italy, domenico.desalvio2@unibo.it

(3)STAT, Department of Statistical Science, University of Bologna, Italy, laura.anderlucci@unibo.it

(4)DIN, Department of Industrial Engineering, University of Bologna, Italy, massimo.garai@unibo.it

Abstract

Speech intelligibility plays a key role in determining the quality of verbal communication. It depends on the acoustic characteristics of the room and the signal-to-noise ratio (SNR). Background noise has three main components: HVAC noise, anthropic noise (student activity) and external activities noise, each of them with different noise spectra. In this work, measurements taken with a sound level meter during lectures in university classrooms are analysed using advanced statistical techniques in order to detect and separate the various components which contribute to background noise. The same statistical techniques are used to characterize the speech level too, i.e. the signal in the SNR. From the raw data collected with a sound level meter, an asymmetrical distribution is built. Then four statistical techniques are applied: percentile levels, Gaussian mixture model based on peak detection, blind Gaussian mixture model and blind k-means clustering. Results are compared and discussed, highlighting the pros and cons of each technique.

Keywords: Speech intelligibility, Gaussian Mixture Model, Signal-to-noise ratio

1 INTRODUCTION

Good quality of oral communication is the crucial aspect for a learning space. A high speech intelligibility allows increasing student's concentration and decreasing teacher's vocal effort [1, 2]. Conversely, a high background noise triggers the *Lombard effect* [8], the mechanism for which a person tends to increase the level of his voice in order to listen himself. As all people experience the same effect, the signal-to-noise ratio (SNR) remains almost constant during time [9]. The main factors affecting the speech intelligibility are the reverberation time and the signal-to-noise ratio. Thus, in order to achieve an high interaction between students and teachers, a deep knowledge of the sound sources affecting the background noise is essential. The best intelligibility in a certain space is reached when the signal-to-noise ratio is at least +15 dB [3, 4]. Often, lecture halls are provided of a public address (PA) system to increase the SNR, but it changes the human noise behaviour too; Peng et al. note that in the case of elementary school classrooms the correlation between SNR and the *Lombard effect* is missed [10]. Background noise in a university lecture hall in occupied condition is formed by HVAC system, external noise due to the nearby spaces and student activity (anthropic noise) [6]. Separating these background noise sources is not immediate when anthropic noise is considered; for example it may be necessary to measure the sound pressure levels during the pauses in the teacher's speech [5]. Statistical techniques offer a viable method to separate the different contributions. However, the standard technique, based on the percentile levels, doesn't allow to obtain the anthropic component in a stable way since it looks for the portion of time exceeded by a random phenomenon. That is why a statistical algorithm is needed to group sound pressure levels recorded with a sound level meter in a more structured way. In this work four statistical techniques are compared to identify the anthropic noise due to students during thirteen university lectures in three different rooms: percentile levels (PL), Gaussian Mixture based on peak detection (PD), blind Gaussian Mixture Model (GMM) and blind k-means clustering (KC). Results are compared and discussed, highlighting the pros and cons of each technique.

2 METHOD

The case studies are three university rooms in the School of Literature and Philosophy of the University of Bologna. Hall I and Hall II have an amphitheater shape with plaster walls and wooden seats and benches, volume of 1000 and 900 m^3 and a maximum capacity of 250 and 200 students. Hall III has a shoe-box shape, plaster walls and movable plastic chairs, a volume of 850 m^3 and a maximum capacity of 170 students.

At first, in order to characterize the acoustics of the rooms, a measurements campaign according to ISO 3382 [11] and IEC 60268-16 [12] was carried out, using a calibrated sound source [14] with enough signal-to-noise ratio [15]. Reverberation time, STI (Speech Transmission Index) and C_{50} (Sound clarity) are considered the relevant acoustic criteria for this study [16].

Subsequently, thirteen lessons were entirely recorded with two sound level meters located at mid distance between the speaker and the rear wall, near the lateral walls of the rooms and at least 1 m from any reflecting surface. An observer detected the unforeseen extraneous sounds in order to delete their peaks in post-processing. Time history were cut in correspondence of the pauses, focusing the analysis on the lectures. Short-term, A-weighted sound pressure levels were recorded with an interval time of a 100 ms. The HVAC system was switched off during the acquisitions.Therefore, only two sound sources were present, labelled in the following as Speech Level (SL) and Student Activity (SA).

The workflow of the post-processing was the following:

1 - the statistical distribution of A-weighted sound pressure levels was built and plotted;

Then, basing on the type of analyses needed:

2.1 - the percentile levels were extracted;

or

2.2 - the distribution of occurrences was fitted by two normal-distribution curves or two clusters, basing on the used technique, to identify each source (SL and SA);

3 - SNR is evaluated as the difference between the local maxima of SL and SA.

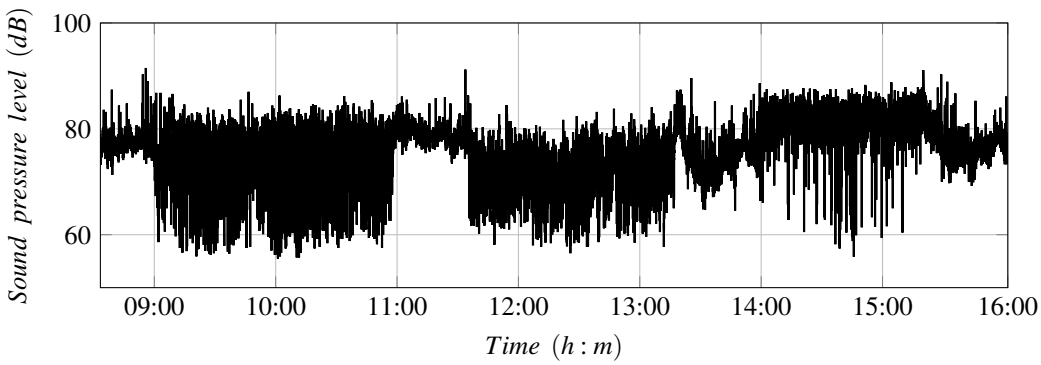


Figure 1. Example: temporal history of three recorded lectures.

If the number of sound sources is known it is possible to "force" the number of clusters assembled by the algorithms or let them to optimize it.

The techniques employed in this work can be divided in two types: with or without an observer during the acquisition. The percentile levels and peak detection are techniques of the first kind. In fact, here the observer

decides the percentile value or the number of geometrical peaks on the occurrences curve to take int account (the observer decides which extraneous noises to discard).

The blind Gaussian Mixture Model and k-means clustering are techniques of the second kind. In fact, they work in a blind way and the operator can just select the number of clusters to obtain. In this study, for the blind techniques a number of 2 clusters was set up.

Table 1. Subdivision between statistical techniques and observer requirement.

Technique	With/without observer
Percentile levels (PL)	with
Peak detection (PD)	with
Gaussian Mixture Model (GMM)	without
K-means clustering (KC)	without

Table 2. Occupancy during lectures (N), gender of the teacher and recording time for each lesson under study.

Lesson	N	Hall	Teacher's gender	Recording time (h:m)
A	145	I	M	01:16
B	200	I	M	01:22
C	100	I	F	01:38
D	150	I	M	01:24
E	250	II	M	01:32
F	160	II	F	01:24
G	120	II	F	01:19
H	150	II	F	01:30
I	200	II	M	01:29
J	110	III	F	01:35
K	80	III	F	01:30
L	110	III	M	01:35
M	175	III	F	01:27

2.1 Percentile Levels

In acoustics a percentile level corresponds to a sound pressure level which is exceeded a certain percentage of the recording time. In this work, using this technique the teacher's speech level is identified with the $L_{A,eq}$ of the whole time history while the student activity is associated to the L_{90}, as in many previous studies [5, 7, 18]. Thus, the SNR is evaluated with their difference. The "acoustic" percentile level corresponds to the rank R of a percentile in statistical terms (see 2.2).

2.2 Peak detection

Peak detection executes a Gaussian Mixture Model algorithm forcing manually the number of possible clusters. This is established by the operator identifying the peaks on the occurrences curve of the recorded SPL. A multi-peak analysis was made with *OriginPro* [13] software in order to identify the main starting points for the Gaussian fitting. The Gaussian Mixture Model algorithm is analysed in detail in the following subparagraph.

The same dataset can be seen in two different ways: through the cumulative distribution or through the occurrence curve (see fig 3). So percentile levels and peak detection techniques are linked. In fact, evaluating a percentile level means doing a backwards integration of the occurrence curve (see figure 3(b)) covering the

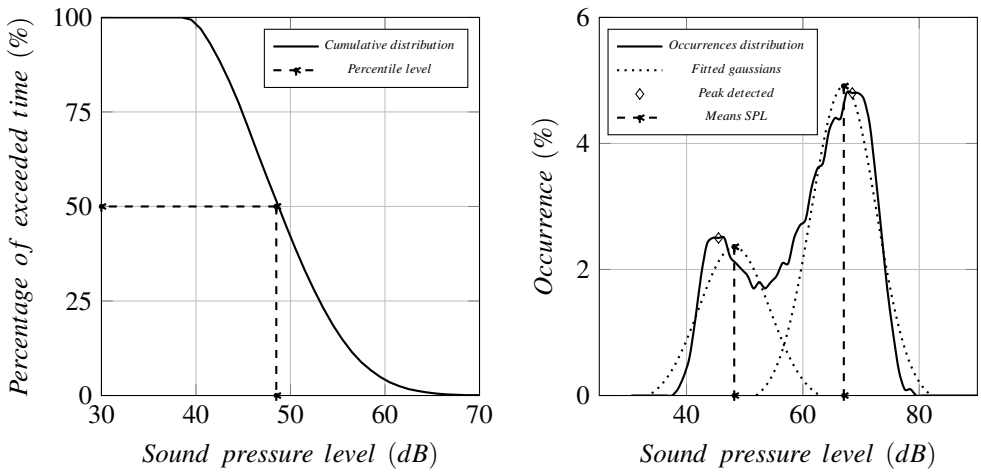
percentage of its area until the percentile required. For example, determining the acoustic percentile level L_{90} corresponds to the backwards integral until the *10*th percentile of the occurrence curve.

In statistics, the rank r of a percentile q of N observations is defined as:

$$r(q) = \frac{q}{100}(N+1) \tag{1}$$

Consequently, the value of the *acoustic* percentile level L_q of a certain dataset is equal to the rank r of 100-q. For a large number of observation, if the the occurrence curve is represented by $f(x)$, the value q can be expressed as:

$$q = P(x > L_q) = \int_{r(100-q)}^{\infty} f(x)dx \tag{2}$$



(a) Percentile Levels technique

(b) Peak Detection technique

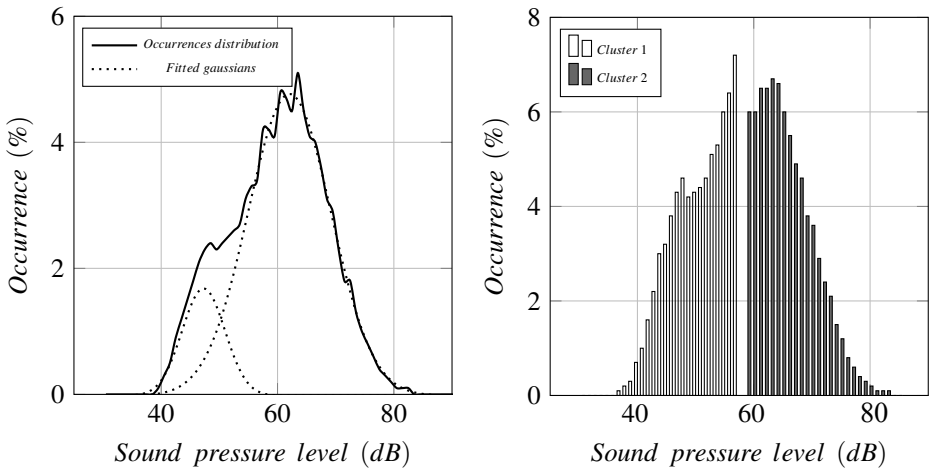
Figure 2. Example of the two techniques with observer. In figure (a) the continuous line represent the cumulative distribution of the SPL recorded by a sound level meter. The example highlights the L_{50} percentile level, which corresponds to an SPL of 48.5 dB. In figure (b) the continuous line represents the distribution of occurrences of the same measurement. The asymmetrical distribution was decomposed in two gaussian curves basing on the peak detection (peaks are indicated with $\diamond$). The means of gaussian curves indicated with $*$ correspond to different sound sources.

2.3 Gaussian Mixture Model

The Gaussian Mixture Model decomposes the original model data in a sum of gaussian curves. The clusters, represented by the gaussian curves, and their points are defined via Maximum Likelihood method. With *Rstudio* [23] is possible estimate via Expectation - Maximization (EM) [22] iterative algorithm the Maximum Likelihood of the recorded SPL.

Let X be a set of independent observations, $x_1, \ldots, x_n$, drawn from a mixture of Gaussian distributions; the density $f(x_i)$ can be written in the form

$$f(x_i; \psi) = \sum_{k=1}^{K} \pi_k f_k(x_i; \theta_k) \quad i = 1, \ldots, n, \quad \psi = \{\theta, \pi\} \tag{3}$$



(a) Blind Gaussian Mixture Model

(b) Blind k-means clustering

Figure 3. Example of the two techniques without observer. In figure (a) the continuous line represents the distribution of occurrences of the recorded SPL while the dashed lines represent the gaussian curves found. In figure (b) are highlighted two clusters of the recorded dataset.

where the $f_k(x_i,\theta)$s are the Gaussian densities with parameter vector $\theta_k = \{\mu_k, \sigma_k^2; k = 1\ldots,K\}$ and π_k are the so called *mixing proportions*, non-negative quantities that sum to one; that is, $0 \leq \pi_k \leq 1 \quad (k = 1,\ldots,K)$ and $\sum_{k=1}^{K} \pi_k = 1$ [21]. The likelihood function for a mixture model with K univariate Normal components is:

$$\mathscr{L}(\psi|x) = \prod_{i=1}^{n}\sum_{k=1}^{K} \pi_k f_k(x_i|\theta_k) = \prod_{i=1}^{n}\sum_{k=1}^{K} \pi_k \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu_k)^2}{2\sigma_k^2}} \tag{4}$$

2.4 K-means clustering

K-means clustering is an algorithm which distributes the data basing on their characteristics, generally the metric used is the distance. As the EM, it is iterative too and identify each cluster with a significant parameter named centroid. Its aim is minimize the intra-cluster variance arranging the centroids [19]. With *Rstudio* [23] it is possible to run the algorithm.

Let $X = \{x_i\}$, i=1, …,n be the set of n points to be clustered into a set of K clusters, $C = \{c_k; k = 1,\ldots,K\}$. K-means algorithm finds a partition such that the squared Euclidean distance between the empirical mean of a cluster and the points in the cluster is minimized. Let μ_k be the mean of cluster c_k. The squared Euclidean distance between μ_k and the points in cluster c_k is defined as

$$J(c_k) = \sum_{x_i \in c_k} ||x_i - \mu_k||^2 \tag{5}$$

The goal of K-means is to minimize the sum of the squared Euclidean distance over all K clusters, therefore the objective function to be minimize is the following:

$$J(C) = \sum_{k=1}^{K}\sum_{x_i \in c_k} ||x_i - \mu_k||^2 \tag{6}$$

3 PRELIMINARY RESULTS

The recorded data were analysed with four statistical techniques. Results are presented in table 3. For each method student activity (SA), speech level (SL) and signal-to-noise ratio (SNR) are shown.

Table 3. Measured A-weighted values of student activity (SA), received speech level (SL) and corresponding signal-to-noise ratio (SNR) for Peak detection, Percentile levels, Gaussian mixture and K-means clustering methods. Values are averaged over the two receiver positions selected for the measurements performed during lessons. In the Equivalent levels method L_{A90} and L_{Aeq} represent the student activity and the speech level. Mean values don't evaluate lesson L since was conducted without P.A. system. All values are in dB.

Lesson	PD			PL			GMM			KC		
	SA	SL	SNR	$L_{A,90}$	$L_{A,eq}$	SNR	SA	SL	SNR	SA	SL	SNR
A	48.0	65.1	17.1	48.0	69.9	22.0	48.2	65.0	16.8	52.2	68.3	16.1
B	47.4	63.5	16.1	45.8	64.8	19.0	47.5	63.3	15.8	48.8	64.2	15.4
C	51.1	65.8	14.7	53.0	68.9	15.9	53.3	66.3	12.9	55.8	68.4	12.6
D	50.8	67.4	16.6	49.3	69.7	20.4	51.2	67.2	16.0	52.7	68.4	15.7
E	47.9	68.2	18.9	52.1	72.3	20.2	48.4	67.5	19.1	49.1	68.0	18.9
F	50.1	66.7	16.7	55.0	71.9	17.0	50.3	66.5	16.2	53.1	68.5	15.5
G	62.0	75.8	13.2	61.6	78.6	17.0	61.0	75.5	14.5	55.7	74.9	19.3
H	55.5	76.1	20.2	56.4	79.2	22.8	55.3	75.3	20.0	55.8	76.0	20.2
I	53.6	68.1	12.9	58.8	74.6	15.8	53.4	68.0	14.6	56.5	69.7	13.1
J	46.9	59.9	13.0	50.3	63.3	13.0	53.0	61.6	8.6	53.3	63.6	10.3
K	50.4	68.2	17.8	47.5	67.8	20.3	50.6	67.6	17.0	50.6	67.7	17.1
L	–	60.7	–	53.4	63.6	10.2	57.7	62.9	5.2	55.8	64.4	8.6
M	48.8	62.7	13.9	51.5	65.1	13.7	51.1	63.1	12.0	53.8	64.7	10.9
Mean	51.0	67.3	15.9	52.4	70.5	18.1	51.9	67.2	15.3	53.1	68.5	15.4

Results show that the differences between the methods decrease when the SNR increases; indeed the lesson "L" highlights this behaviour. This lesson was carried out without the PA system so, because of the *Lombard effect*, the student activity is not visually appreciable. Thus, the PD technique doesn't succeed in the fitting while other techniques give very different results. The density of the data doesn't permit to have a coherent clustering. The most similar values are obtained by the PD and GMM, since they use the same algorithm but in a different initial conditions. More in general, PL overestimates SNR values compared to the other techniques.

Measurements give further insights into the mechanism of the *Lombard effect*. Students and teachers create an acoustical environment with an SNR of +15 dB as a consequence of the *Lombard effect* to reach the maximum possible speech intelligibility in the rooms as indicated by previous studies [17, 9]. As shown in figure 4 *SL* increases of 0.56 dB for each dB of increase of *SA* as suggested by literature [24, 25]. The intercept (37.4) takes into account the combined effect of the source-receiver distance and the PA gain.

4 CONCLUSIONS

The present work has investigated a statistical analysis of the acoustic field during university lessons. After a measurement campaign carried out with two sound level meters, statistical analysis were conducted with four different techniques: percentile levels, Gaussian Mixture model based on peak detection, blind Gaussian Mixture Model and blind k-means clustering. Results show how the percentile levels overestimate, unlike the other techniques, the value of the background noise during university lectures. Furthermore the reciprocity between the student activity and the speech level highlights how *Lombard effect* works maintaining a costant SNR.

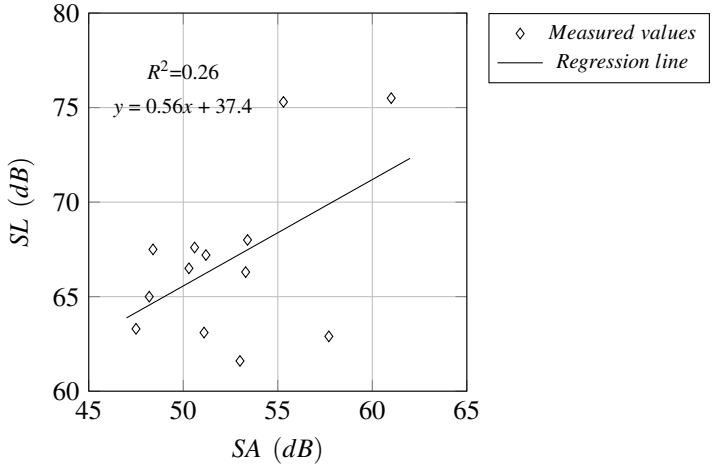


Figure 4. Relationship between *SA* and *SL*. The regression line shows the increase of 0.56 dB/dB as suggested by literature.

REFERENCES

[1] Visentin C., Prodi N., Cappelletti F., Torresin S., Gasparella A., Using listening effort assessment in the acoustical design of rooms for speech, Build. Environ., 136 (2018), 38–53.

[2] Puglisi G. E., Cutiva L. C., Pavese L., Castellana A., Bona M., Fasolis S., Astolfi A., Acoustic comfort in high-school classrooms for students and teachers. Energy Procedia, 78 (2015), 3096–3101.

[3] Reich R., Bradley J., Optimizing classroom acoustics using computer model studies, Canadian Acoustics, 26(4) (1998), 15–21.

[4] Hodgson M., Experimental investigation of the acoustical characteristics of university classrooms, J. Acoust. Soc. Am., 106(4) (1999), 1810–1819.

[5] Bottalico P., Astolfi A. (2012). Investigations into vocal doses and parameters pertaining to primary school teachers in classrooms. The Journal of the Acoustical Society of America, 131(4), 2817-2827.

[6] Hodgson M., Rempel R., Kennedy S., Measurement and prediction of typical speech and background noise, J. Acoust. Soc. Am., 105(1) (1999), 226–233.

[7] Shield B., Conetta R., Dockrell J., Connolly D., Cox T., Mydlarz C., A survey of acoustic conditions and noise levels in secondary school classrooms in England, J. Acoust. Soc. Am., 137(1) (2015), 177–188.

[8] Lombard E., Le signe de l'elevation de la voix, Ann. Mal. de L' Oreille et du Larynx XXXVII (1911), 101-119.

[9] Sato H., Bradley J., Evaluation of acoustical conditions for speech communication in working elementary school classrooms, J. Acoust. Soc. Am., 123(4) (2008), 2064–2077.

[10] Peng J., Zhang H., Wang D., Measurement and analysis of teaching and background noise level in classrooms of Chinese elementary schools, Applied Acoustics, 131 (2018), 1-4.

[11] ISO 3382-2:2008, Acoustics - Measurement of room acoustic parameters - Part 2: Reverberation time in ordinary rooms. Geneva: International Organization for Standardization.

[12] IEC 60268-16:2011, Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index. Geneva: International Electro-technical Commission.

[13] OriginPro 2017. OriginLab Corporation, 2016.

[14] D'Orazio, D., De Cesaris, S., Guidorzi, P., Barbaresi, L., Garai, M., Magalotti, R. Room acoustic measurements using a high-SPL dodecahedron (2016) 140th Audio Engineering Society International Convention 2016, AES 2016,

[15] P. Guidorzi, L. Barbaresi, D. D'Orazio, M. Garai, Impulse responses measured with MLS or Swept- Sine signals applied to architectural acoustics: an in-depth analysis of the two methods and some case studies of measurements inside theaters, Energy Procedia, 78 (2015) 1611–1616.

[16] De Cesaris, S., F. Morandi, L. Loreti, D. D'Orazio, and M. Garai (2015). Notes about the early to late transition in italian theatres. Proc. of ICSV22, Florence .

[17] Hodgson M., Nosal E., Effect of noise and occupancy on optimal reverberation times for speech intelligibility in classrooms, J. Acoust. Soc. Am., 111(2) (2002), 931–939.

[18] Titze, I. R., Hunter, E. J., Å vec, J. G. (2007). Voicing and silence periods in daily and weekly vocalizations of teachers. The Journal of the Acoustical Society of America, 121(1), 469-478.

[19] Jain, A. K. (2010). Data clustering: 50 years beyond k-means. Pattern Recognition Letters, 31(8), 651–666.

[20] Lloyd, S. (1982). Least squares quantization in pcm. IEEE Trans. Inf. Theor. 28 (2), 129–137.

[21] McLachlan, G. J. and D. Peel (2000). Finite Mixture Models. Wiley.

[22] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), Maximum Likelihood for Incomplete Data via the EM Algorithm (with discussion). Journal of the Royal Statistical Society, Ser. B, 39, 1–38.

[23] R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2017.

[24] Rindel, J. H. (2010). Verbal communication and noise in eating establishments. Applied Acoustics, 71, 1156–1161.

[25] Lazarus, H. (1986). Prediction of verbal communication in noise–a review: part 1. Applied Acoustics, 19, 439–464.