AUTHOR QUERY FORM

AIP Publishing	Journal: J. Acoust. Soc. Am. Article Number: 022510JAS	Please provide your responses and any corrections by annotating this PDF and uploading it according to the instructions provided in the proof notification email.

Dear Author,

Below are the queries associated with your article; please answer all of these queries before sending the proof back to AIP.
Please indicate the following:
Figures that are to appear as color online only (i.e., Figs. 1, 2, 3) ______________________________(this is a free service).
Figures that are to appear as color online and color in print______________________(a fee of $325 per figure will apply).

Article checklist: In order to ensure greater accuracy, please check the following and make all necessary corrections before returning your proof.

1. Is the title of your article accurate and spelled correctly?
2. Please check affiliations including spelling, completeness, and correct linking to authors.
3. Did you remember to include acknowledgment of funding, if required, and is it accurate?

Location in article	Query / Remark: click on the Q link to navigate to the appropriate spot in the proof. There, insert your comments as a PDF annotation.
AQ1	Please check that the author names are in the proper order and spelled correctly. Also, please ensure that each author's given and surnames have been correctly identified (given names are highlighted in red and surnames appear in blue).
AQ2	Please define "CVCV" at first occurrence.
AQ3	Please confirm the table caption as edited.
AQ4	Please clarify what is meant by "darker grey background" and "lighter one." Is this a reference to the bolding?
AQ5	Correct as changed to "Table IX"?
AQ6	Please provide a digital object identifier (doi) for Ref(s). 2,4,26, and 29. For additional information on doi's please select this link: http://www.doi.org/. If a doi is not available, no other information is needed from you.
AQ7	Please provide page range for Ref(s). 4 and 15.

Thank you for your assistance.

Listening efficiency during lessons under various types of noise

AQ1

Nicola Prodi[a)] and Chiara Visentin

Dipartimento di Ingegneria, Università di Ferrara, Ferrara, Italy

(Received 22 June 2015; revised 14 September 2015; accepted 18 September 2015; published online xx xx xxxx)

Pupils inside primary school classrooms can be exposed to extraneous noise, impairing their performance in the speech reception process. The different noises show a peculiar impact, depending on their level, spectral content and temporal fine structure. In order to understand how the disturbance is built up over time, in this work a large data set was analyzed, detailing the changes of the pupils' performance as the lesson progresses from the start to the end. Several types of noise are considered (traffic, tapping and babble noise) and the analysis concerns III to V graders of the Italian primary school (8 to 10 year old pupils). By using as indicators the intelligibility scores, the response time and their ratio, the so-called "listening efficiency," several findings are achieved. Pupils respond differently to each noise during the course of the lesson. In the best listening conditions, the performance in the speech reception worsens under traffic and babble noise whereas an opposite trend is found under tapping noise. On the contrary adaptation is observed in the worst listening conditions for the traffic noise alone. Moreover, indications are achieved that the age proficiency may affect differently babble noise compared to traffic and tapping noise.

 [http://dx.doi.org/10.1121/1.4932053]

[NX] Pages: 1–11

I. INTRODUCTION

The effect of noise on pupils' performance at school has been widely investigated during the years, showing that classroom acoustics affects children listening, learning and behavior.[1–4] Chronic exposure to high environmental noise leads to long-term effects: the analysis of the children scores on accuracy tasks (e.g., word recognition or standardized academic tests) highlights the presence of detrimental effects on memory and reading ability and a reduction of children's attention and motivation.[5] In addition, the presence of noise inside classrooms requires the pupils to pay more attention to the speech recognition process, increasing the time and the effort to process the information.[6–8] Therefore, when lessons are held inside noisy classrooms, the pupils continuously exert themselves to understand the teacher. The most critical conditions are manifested with structured lexical tasks, such as text comprehension or item retention in the short-term memory, even when the intelligibility of words is nearly perfect.[9,10] Thus, only when speech intelligibility is high and at the same time speech reception is easy there can be a release of working memory resources for elaboration, recording, storing and subsequent recall of information which is typical of a learning process.[11]

In order to deal with classroom acoustics the most common approach among regulations is based on granting a certain speech intelligibility and a limited noise annoyance by prescribing appropriate reverberation and sufficiently low noise level, and is implemented by comparing measurable indicators with their ranges of suitability. Despite the robustness and widespread usage of the above concepts, they may be not entirely reliable to describe the impact that bad acoustics has on the basic elements (for instance memory usage) that build up the learning process.[12,13]

On the other hand, only tests on a range of specific tasks involving several relevant cognitive functions could be appropriate to quantify the impact of acoustics on learning, but such procedures are impractical as normative tools, and at the moment, their usage is restricted to research purposes. Thus, it appears that, besides accuracy in speech reception brought by intelligibility scores, additional information on how "easy" (or "effortless" or "not difficult") listening is in the classroom could be a feasible solution in order to enhance the current means of qualification with a measure that tapers a prerequisite for effective learning.

"Listening effort" has been the subject of several studies in the field of audiology and can be assessed by dual-task experiments, by means of subjective scales of various type and by using diverse physiological indices whose rationale is resumed in Ref. 14. Recently, the usage of "response time," that is the time elapsed from target item offset to response onset, was further validated as a simple "listening effort" measure with speech sentences.[15] Moreover a rough estimate of "listening effort" was also obtained for adults with *speech transmission index* (STI) measures.[16,17] Similarly to the subjective scaling proposed for "listening effort," the "listening difficulty"[18] consists in the reporting of subjective impressions on a scale of four items. Previous studies have shown that, under favorable signal-to-noise ratios, "listening difficulty" could be more effective than intelligibility scores in discerning differences between the listening conditions.[18] When applied to children, "listening difficulty" showed some inconsistency,[8] which had been pointed out for pupils also in Refs. 6 and 19 and is due to a possible mismatch between the subjective impression and the effective objective performance.

[a)]Electronic mail: nicola.prodi@unife.it

To avoid any ambiguity in the qualification of effort, an approach excluding subjective impressions was proposed, and was also motivated with reference to basic cognitive mechanisms.[8] The method involves intelligibility scores (IS) as accuracy measures and the "response time" (RT), which as in Ref. 15 is the time elapsed from the end of item presentation to the subject's response, as a psychophysical measure of effort.

The ratio of IS and RT is termed "listening efficiency," is referred to by DE since it is the "direct" ratio of the quantities,[8] and is nominally the number of correct items spot within one second. DE has proved to be able to much better identify pupils' performance in speech reception under adverse acoustical conditions in classrooms rather then IS alone.[8] The present study investigates how the "listening efficiency" together with speech intelligibility and response time of III, IV, and V grade pupils (aged 8, 9, and 10, respectively) behave during a time interval equivalent to a lesson period. In particular, it is researched whether the performance is stable or not depending on the type of noise and on the proficiency of the students. This information is valuable in order to better understand the specific way noise hampers the overall cognitive features during the lesson by exerting and overloading the pupils with extra effort to reach a given accuracy. In fact, the knowledge of such mechanisms is needed to foster future normative limits and to highlight priorities that are to be considered when coping for the noise, in particular during the acoustical design of the school buildings.

The present study is based on a subset of the data collected in Ref. 20, which are rearranged for the present purposes as described in what follows. The materials and methods are explained in Sec. II and in Sec. III, respectively. Results are presented in Secs. IV and V, and discussed in Sec. VI.

II. OVERVIEW OF DATA SET

A. *In situ* measures

The experiments took place in seven parallelepiped classrooms which served as laboratories inside six primary schools in the city of Ferrara, Italy. None of the rooms had an acoustical treatment and just the interior furniture contributed to the sound absorption with desks, chairs, few maps and seldom book shelves. The classroom volumes span from 121 m^3 to 187 m^3 and the unoccupied mid frequency (average of 500 Hz and 1 kHz octave bands) reverberation times T_M varied between 1.65 and 0.90 s. To enlarge the ensemble of conditions each room was also temporarily equipped, during half of the tests, with sound absorbing melamine blankets in order to further decrease the reverberation time. The range of T_M in occupied conditions and with or without temporary treatment varied between 1.00 and 0.60 s.

B. Tests setup

Tests here considered are for the III, IV, and V grade (that is from 8 to 10 year old pupils), whose testing material consisted in a Diagnostic Rhyme Test (DRT) in the Italian language.[21] This material is phonetically balanced and consists in pairs of CVCV rhyming words.

Three types of noises were selected to interfere with the speech signal, called, respectively, "babble and activity" (A), "tapping" (T_p), and "traffic" (T_r). The first one was a continuously fluctuating signal created by processing Italian audiological test phrases[22] according to the established ICRA instructions.[23] The processed signal has the same frequency and temporal spectral characteristics as the natural speech while carrying no semantic meaning and being completely unintelligible. To this signal a few typical activity noises were added by digital mixing such as rolling of a pen, falling of a pen and turning over of book pages. The tapping noise (T_p) was obtained by recording inside a silent laboratory room while impact noise was generated on the floor upstairs. This noise was due to the dragging of several chairs and hitting the floor with a pole. Finally, the traffic noise (T_r) was recorded on the side of a busy road a few meters from the track. The long-term averaged spectra of the noises are shown in Ref. 20, together with their temporal structure. During the test session, A and T_p were played back with an omnidirectional source placed inside the classroom whereas for T_r a sound system directional loudspeaker was used, placed outside the school building, two meters away from the facade and directed toward the classroom windows.

The speech material, consisting in the target words preceded by a carrier phrase, was read by a native female speaker and was recorded in a silent room with an omnidirectional microphone at a sampling rate of 44.1 kHz. The signal was played back inside the classrooms by a directional loudspeaker at "raised," "loud," or "normal" vocal efforts,[24] measured at 1 m in front of the loudspeaker. The noise levels were varied accordingly as to obtain, at the same position, sound to noise ratios (SNRs) of 0, 6, and 12 dBA. In particular, for the "babble and activity" condition, the vocal effort was fixed at 66 dBA and the noise level changed, whereas for traffic and tapping the noise level was fixed at 60 dBA and the vocal effort was adjusted.

The duration of the experiment for each class was 45 to 55 min, which is comparable to the duration of a lesson. In the planning of the tests that interval was subdivided into nine slots of almost 5 to 6 min each, corresponding to the time needed to the children to go through a basic test unit composed of seven DRT word pairs. Each slot of time was characterized by one background noise (T_r, T_p, or A) whereas the vocal effort varied between the seven words. In the planning of the tests the nine conditions investigated (3 noises $\times$ 3 vocal efforts) had been assigned across the slots and the classes following a Latin-square design, so that both systematic learning effects were avoided and equal probability of the test conditions across the slots was ensured.

C. Data collection

The test was administered to all pupils after written parental consensus. The initial number of pupils was 589 and 530 of them were validated. Exclusion of participants was evaluated with the help of teachers based on specific learning disorders or listening deficits. Pupils were distributed in 33 classes; the population in each class ranged from 16 to 24 (Table I). The tests were administered with an automated system which is capable of managing at once the audio playback and the collection of the responses. Details on the system (called *Intelligo*) can be found elsewhere.[25] Here it is to be noted that every pupil had a touchscreen mobile phone in the hands and the words to choose appeared as screenshots immediately after the target word with noise was proposed (teacher's voice with noise). By these means it was possible to record response time RT, which is the time elapsed from the offset of the target stimulus to the instant when the selection on the touchscreen was done. Once the IS are calculated also DE can be obtained as the ratio of IS and RT.

The objective description of the conditions during the listening tests was carried out with the monaural speech transmission index STI: after the experiment, for each of the 33 classes, impulse responses were collected and signal and noise levels measured at four measurement positions. In the following analysis, a four-position averaged value of STI was used to characterize the whole classroom space under a given condition. The choice was supported by the limited variations of the STI values in the classrooms, resulting in a maximum deviation from the average value of ±0.05 (found in the largest location with the less numerous class). However, in 90% of the classes the deviation was limited to 0.04, which is a reference interval derived in Ref. 16 from the typical uncertainty of STI values.

III. RUN-TIME ANALYSIS

In the present work the changes of DE during a time period typical of a lesson are analyzed: the results for a given noise which were collected from different classes of equal grade are used to fill the nine slots, in order to derive a sort of time-history of the test results for that case. An example of the procedure employed is reported in Table II for two III grade classes. Each "slotted" point represents the average over the four measurements positions which hosted all the pupils of a class. Because of the randomization of the tests, some slots had more points than others while some slots had only few points for some conditions.

In order to prepare the data for statistical analysis, they were further stratified for each grade and noise condition according to the STI ranges that in the norm[24] are matched with the ratings of communication quality, that is $STI < 0.30$ for *Bad*, $0.30 < STI < 0.45$ for *Poor*, $0.45 < STI < 0.60$ for *Fair*, and finally $0.60 < STI < 0.75$ for *Good*. Data were further merged in a *Good* + *Fair* stratum on one side and in a *Bad* + *Poor* on the opposite side of the STI range.

TABLE I. Number of classes and the related number of pupils for each grade participating in the experiments.

			Population			
Grade	Age	Number of classes	All	Valid	M	F
III	8	11	197	184	89	95
IV	9	13	214	189	101	88
V	10	9	178	157	74	83
Total		33	589	530	264	266

TABLE II. Example of the derivation of run-time data using two III grade classes. The series of noises and conditions for each class was obtained with a latin-square design. Then the slot sequences for the noises (A, T_p, or T_r) were obtained by filling each slot with the data coming from those classes which performed the specific noise test during the slot under consideration. For example in the table the third slot for T_r will include data from class IIIA, and that for T_p will include data for IIIB and so forth.

	Test Sequence: Slot Number								
	1	2	3	4	5	6	7	8	9
IIIB	A	T_r	T_p	T_r	A	T_r	T_p	A	T_p
IIIA	T_r	T_p	T_r	A	T_p	A	T_r	T_p	A
Slotted Noise Sequence									
A	IIIB			IIIA	IIIB	IIIA		IIIB	IIIA
T_r	IIIA	IIIB	IIIA	IIIB		IIIB	IIIA		
T_p		IIIA	IIIB		IIIA		IIIB	IIIA	IIIB

Finally, after stratification, the population of the slots for each stratum could be fixed and is reported in Table III. As can be seen the *Good* stratum in the A noise presents limited population since only very few classrooms actually had favorable conditions. In this and similar cases the *Good* stratum was not analyzed independently and only the pooled *Good* + *Fair* intervals were considered.

It is to be remarked that the present analysis differs substantially from the data arrangement in Ref. 20. In fact in Ref. 20, the DE data were pooled by grade and noise disregarding the specific slot, so that the results achieved there could be only considered as "lesson time-averages," without any insight into the trend of quantities during the course of the time interval typical of a lesson. On the contrary the present elaborations will deal exclusively with the slotted data in order to investigate how the lesson-averaged values are built up by the run-time trends. Since DE is obtained as the ratio of IS and RT it is useful to apply the same analysis to the latter quantities as described in what follows.

IV. RESULTS

A. Overview of significant cases

The course of the performance, effort, and speech intelligibility during the lesson period can be evaluated

TABLE III. Number of samples for the strata. For each grade and stratum the number of samples in the cell corresponds to the sum over the nine slots of that condition (i.e., for III grade, T_r noise and "*Bad* + *Poor*" stratum the 50 samples are spread over the nine slots). Each single sample is obtained as the average of four positions in one classroom. Each position hosted from four to six pupils.

		Bad + Poor	Bad	Poor	Fair	Good	Fair + Good
III	T_r	50	16	34	29	6	35
	T_p	49	17	32	38	9	47
	A	20	11	9	9	2	11
IV	T_r	67	26	41	35	9	44
	T_p	60	21	39	45	12	57
	A	26	12	14	11	1	13
V	T_r	43	19	24	33	6	39
	T_p	39	12	27	32	9	41
	A	15	7	8	8	2	10

graphically and, as an example, the course of DE for the A, T_r, and T_p noises is shown in Fig. 1 for grade V. In the plots the set of data in each slot for a given stratum have been averaged to allow a more clear presentation (e.g., in Fig. 1 the point corresponding to the first slot of the *Fair* stratum is the average DE value of the six classes performing the test in that acoustic condition). The best fit linear regressions are also included for ease of trend recognition. As it can be seen from Fig. 1, a general course of the values can be graphically outlined in some cases and the same would be true for few trends in other grades, noises, and quantities.

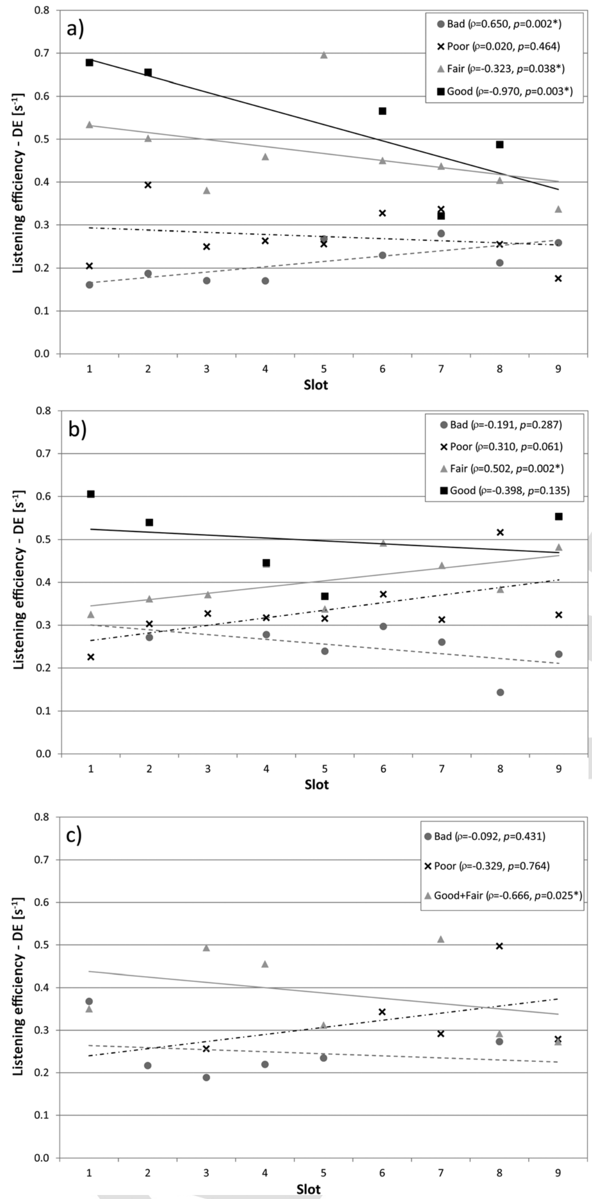

FIG. 1. Plot of listening efficiency (DE) regressions for the V grade. In (a) the T_r noise, in (b) the T_p noise and in (c) the A noise. In order to have a sufficient ease of reading in the figures, a simple modification of the data was accomplished. Each point in the plots is the average of the values of the classes concurring at that slot. For the same scope the best fit linear regression lines are drawn on the data for each case. Spearman's correlation ρ coefficients and the respective p values are included. An asterisk indicates significance at the $\alpha = 0.05$ level. In frame (c) the strata *Good* and *Fair* are merged in the presentation as in the analysis (see also Tables V, VI, and VII), since the *Good* stratum has only 2 points (see Table III last line).

In order to validate the trends and better investigate also the effect of speech intelligibility and response time, a dedicated statistical analysis was implemented which was based on a rank ordering correlation of Spearman type between slot number and the quantity under investigation (in turn DE, RT, and IS). Since the points within each stratum had not the same objective STI values, and STI may itself correlate with the three quantities,[8] it was necessary to control for the point-specific STI by employing a partial correlation analysis, where such dependence could be taken out of the main expected effect. So the data were processed by partial correlation and the significance of the results was evaluated too. In this case the null hypothesis (H_0) is that there is no correlation whereas the alternative hypothesis (H_1) is that there is non-zero correlation or, in other words, that the trend increases (or decreases) during the equivalent lesson time. The respective p values were thus calculated by using a specific right or left tail depending on the expected effect. The choice of H_1 in fact was decided after a preliminary statistical analysis performed in Refs. 26 and 27, which first showed some gross trends.

Table IV resumes the number and type of significant cases for the Spearman's partial correlation coefficients at the level of $\alpha = 0.05$. According to the effect size interpretation of correlation coefficients provided in Ref. 28, these data are subdivided between $0.3 < |\rho| < 0.5$ (medium effect size) and $|\rho| \geq 0.5$ (large effect size) while data in the range $0.1 < |\rho| < 0.3$ (small effect size) are not included in Table IV since no such value reached significance at $\alpha = 0.05$. The last row of Table IV reports the totals of the significant cases according to quantity. It is seen that for 5 (IS), 16 (RT), and 10 (DE) out of 48 values (6 strata $\times$ 3 noise types $\times$ 3 grades minus the 6 clustered *Fair* and *Good* strata under A), the null hypothesis H_0 of no change during the lesson period has to be rejected.

Being listening efficiency the ratio of IS and RT, its significant trends develop from a complex interplay of the other two. Only for a subgroup the three quantities are all significant at $\alpha = 0.05$ while in some cases DE satisfies a looser $\alpha = 0.1$: this will happen primarily when some significant RT trends are not confirmed by the respective IS values.

As regards the strength of the correlations in terms of Spearman's ρ, there is a prevalence of medium effect sizes but several large ones are found. In particular RT points out the biggest number of $|\rho| \geq 0.5$ trends (seven cases),

TABLE IV. Number and repartition of the significant Spearman's partial correlation cases at $\alpha = 0.05$ according to the quantities under examination and to pupils' grades. The numbers in parenthesis indicate strength of effect (Ref. 28). Leftmost figure is for $|\rho| \geq 0.5$ (large correlation) while rightmost one indicates $0.3 < |\rho| < 0.5$ (medium correlation). The totals according to quantity are reported in the last row and those related to grade in the last column. AQ3

	IS (%)	RT (s)	DE (s^{-1})	Total
V	4 (1;3)	9 (4;5)	8 (4;4)	21 (9;12)
IV	1 (0;1)	4 (2;2)	2 (0;2)	7 (2;5)
III	—	3 (1;2)	—	3 (1;2)
Total	5 (1;4)	16 (7;9)	10 (4;6)	31 (12;19)

whereas DE (four) and IS (one) are generally not as effective.

The sums of the row values are reported in the rightmost column of Table IV to obtain the totals according to the grade. The figures increase with the age of the testers from 3 (III grade, all obtained by RT) to 7 (IV grade) and finally to the V grade with a remarkable 21.

Despite a satisfactory and useful data set was accomplished, and that the trends could be outlined, the statistical analysis highlighted that significant correlations existed only for a minority of the cases. This was actually expected due to the nature of the slotted data set. In fact the values included within each slot pertain to different classrooms, which belong to different schools and, although they share the same STI stratum, they refer to not coincident specific STI values and to *in situ* acoustical conditions that may differ somehow. By definition, speech transmission index values can be realized with different combinations of spectral signal-to-noise ratio and reverberation time, and the respective perceived subjective impressions may be not entirely matching. This occurrence causes unavoidable variability in the data set but, on the other hand, allows to generalize the present findings to realistic *in situ* conditions without the need for further *a priori* assumptions apart from volume, reverberation time, and type of noise.

B. Better acoustical conditions

In the Tables V, VI, and VII, all of the specific ρ and p values are detailed, respectively, for DE, IS, and RT; p values are in bold character when below $\alpha = 0.05$ and in Italic when below $\alpha = 0.1$. The top frame is for grade V, while the mid and bottom frames are for IV and III grade, respectively.

The subset of significant correlations can be closely investigated, starting from the most favorable acoustical conditions (that is strata *Fair*, *Good*, and *Fair + Good*). It is seen in Table V for DE that the V grade shows significant results for the three noises in the *Good + Fair* stratum and for T_r this happens separately for *Good* ($p = 0.003$) and *Fair* ($p = 0.038$). Activity noise shows the largest effect size with a negative correlation ($\rho = -0.666$) and also T_r presents negative correlation ($\rho = -0.381$), but the effect size is almost halved with respect to A. So the performance of V graders is not stable during the equivalent lesson hour, but in case of A and T_r it may decrease, and statistics indicate that this finding is easier to detect for A. Then T_p noise shows in the *Good + Fair* stratum a significant reversed trend caused by a positive correlation with a medium effect size ($\rho = +0.332$). This behavior is driven by the *Fair* stratum whereas the *Good* one, though not significant, shows a decreasing trend. That is to say that, contrary to the previous cases, performance under T_p can even improve during the lesson. The values for IS and RT corresponding to the above conditions are found in Tables VI and VII. For IS, the values in T_r for *Fair* ($p = 0.006$) and *Good + Fair* ($p = 0.002$) are congruent with DE whereas the *Good* stratum fails slightly significance ($p = 0.08$). Data for both T_p and A are generally not significant. On the contrary, the V grade RT significant cases in Table VII are matched to the respective DE ones almost completely (the only exception is in the T_r *Fair* stratum).

TABLE V. Listening efficiency DE: Spearman's partial correlation coefficients and related p values for III, IV and V grades under T_r, T_p and A noises. The partial correlation controls for the STI values as to take out the dependency between DE and STI. Values in bold are $p < 0.05$ and in italic when $p < 0.1$. Data for "Good" and "Fair" in case of noise A are merged due to smaller sample size of the two strata in isolation (see Table III).

	Traffic (T_r)		Tapping (T_p)		Babble (A)	
	ρ	p Value	ρ	p Value	ρ	p Value
Grade V						
Bad + Poor	0.305	**0.026**	0.215	*0.097*	0.155	0.298
Bad	0.650	**0.002**	−0.191	0.287	−0.092	0.431
Poor	0.020	0.464	0.31	*0.061*	−0.329	0.764
Fair	−0.323	**0.038**	0.502	**0.002**	—	—
Good	−0.970	**0.003**	−0.398	0.165	—	—
Good + Fair	−0.381	**0.01**	0.332	**0.018**	−0.666	**0.025**
Grade IV						
Bad + Poor	−0.124	0.158	−0.107	0.209	0.224	0.141
Bad	−0.313	*0.064*	−0.283	0.113	0.363	0.136
Poor	−0.037	0.41	−0.092	0.291	0.369	0.108
Fair	−0.334	**0.027**	−0.073	0.319	—	—
Good	−0.398	0.165	−0.255	0.225	—	—
Good + Fair	−0.361	**0.009**	−0.112	0.205	0.148	0.332
Grade III						
Bad + Poor	0.017	0.545	−0.005	0.514	0.315	*0.095*
Bad	−0.045	0.437	0.154	0.285	0.545	*0.052*
Poor	0.128	0.139	−0.117	0.532	−0.219	0.602
Fair	0.197	0.168	−0.099	0.28	—	—
Good	−0.127	0.436	−0.234	0.289	—	—
Good + Fair	0.085	0.676	−0.129	0.196	−0.302	0.198

TABLE VI. Speech intelligibility IS: Spearman's partial correlation coefficients and related p values for III, IV and V grades under T_r, T_p and A noises. Values in bold are $p < 0.05$ and in italic when $p < 0.1$.

	Traffic (T_r)		Tapping (T_p)		Babble (A)	
	ρ	p Value	ρ	p Value	ρ	p Value
Grade V						
Bad + Poor	0.060	0.354	−0.074	0.328	0.001	0.49
Bad	0.441	**0.038**	−0.582	**0.030**	−0.407	0.212
Poor	−0.155	0.239	0.069	0.369	−0.228	0.311
Fair	−0.444	**0.006**	0.268	*0.072*	—	—
Good	−0.731	*0.080*	−0.377	0.178	—	—
Good + Fair	−0.462	**0.002**	0.124	0.222	−0.483	*0.094*
Grade IV						
Bad + Poor	−0.220	**0.037**	−0.138	0.149	−0.032	0.440
Bad	−0.335	*0.051*	−0.207	0.191	−0.236	0.250
Poor	−0.162	0.156	−0.135	0.210	0.081	0.397
Fair	−0.259	*0.070*	−0.001	0.496	—	—
Good	0.082	0.424	−0.136	0.345	—	—
Good + Fair	−0.226	*0.073*	−0.049	0.361	0.092	0.394
Grade III						
Bad + Poor	0.022	0.440	−0.012	0.470	−0.175	0.240
Bad	0.070	0.408	0.086	0.376	0.074	0.420
Poor	0.007	0.490	−0.098	0.402	−0.472	0.119
Fair	−0.005	0.490	−0.115	0.249	—	—
Good	−0.409	0.247	−0.049	0.455	—	—
Good + Fair	0.085	0.324	−0.118	0.218	−0.233	0.259

TABLE VII. Response time RT: Spearman's partial correlation coefficients and related p values for III, IV and V grades under T_r, T_p and A noises. Values in bold are $p < 0.05$ and in italic when $p < 0.1$.

	Traffic (T_r)		Tapping (T_p)		Babble (A)	
	ρ	p Value	ρ	p Value	ρ	p Value
Grade V						
Bad + Poor	−0.415	**0.003**	−0.325	**0.023**	−0.331	0.124
Bad	−0.743	**0.001**	0.001	0.499	−0.146	0.609
Poor	−0.279	*0.098*	−0.428	**0.015**	−0.557	0.453
Fair	0.245	*0.092*	−0.528	**0.001**	—	—
Good	0.932	**0.010**	0.308	0.229	—	—
Good + Fair	0.311	**0.031**	−0.435	**0.002**	0.6809	**0.022**
Grade IV						
Bad + Poor	−0.063	0.307	0.005	0.484	−0.454	**0.011**
Bad	0.076	0.641	0.190	0.211	−0.578	**0.031**
Poor	−0.144	0.185	−0.007	0.516	−0.600	**0.015**
Fair	0.276	*0.057*	0.167	0.140	—	—
Good	0.367	0.186	0.197	0.281	—	—
Good + Fair	0.317	***0.019***	0.192	*0.078*	−0.171	0.308
Grade III						
Bad + Poor	−0.113	0.778	−0.097	0.256	−0.403	**0.044**
Bad	0.051	0.428	−0.265	0.161	−0.608	**0.031**
Poor	−0.230	0.103	0.018	*0.077*	−0.060	0.110
Fair	−0.358	**0.036**	0.027	0.438	—	—
Good	0.244	0.346	0.289	0.243	—	—
Good + Fair	0.085	0.324	0.062	0.341	0.276	0.220

Coming to the DE for the IV graders (Table V, central frame), they are subject to decrease in T_r only ($p = 0.009$ in *Good + Fair* and $p = 0.027$ in *Fair*), whereas no trend is depicted for A and T_p. This occurrence is confirmed by the respective response time cells in Table VII only for the *Good + Fair* stratum, whereas IS (Table VI) does not show any significant trend.

Finally, the better acoustical conditions are examined for the III grade. Scanning quantities and noises there appears just one isolated significant value for RT with noise T_r in the *Fair* stratum, but this value is not confirmed neither by the other quantities nor in the other strata, and thus is not discussed further.

Putting together the data presented so far regarding better acoustical conditions, the trend effects are demonstrated and it can also be argued that the phenomena are mediated by the intrinsic nature of noise, which regulates the prevalence or not of one course (increase or decrease) depending on the acoustical conditions.

C. Worse acoustical conditions

The analysis of Table V for DE is to be extended to the worse strata (*Bad*, *Poor*, *Bad + Poor*), where surprisingly some significant rank correlations are found. This happens in the case of V grade for T_r, *Bad + Poor* ($p = 0.026$) and *Bad* ($p = 0.002$) where an increasing trend of DE is observed. It is noteworthy that the same cases are found to be significant for RT (Table VII) and partly for IS (only for *Bad*). Passing to DE for T_p, V grade, in *Bad + Poor* and in *Poor*, a similar increasing tendency is manifested, but both cases fail slightly the significance testing though having not negligible effect size, that is *Bad + Poor:* $p = 0.097$, $\rho = 0.215$ (small effect size) and *Poor:* $p = 0.061$, $\rho = 0.310$ (medium effect size). Differently, in these cases the corresponding cells for RT in Table VII are both significant (*Bad + Poor* with $p = 0.023$ and *Poor* with $p = 0.015$) but this does not happen for IS.

Thus, for V grade, the improvement of performance during the more noisy lessons, which could be referred to as "adaptation," is entirely effective in T_r, whereas in T_p is only partially so, thanks to the significant RT trend. Other graders under T_p and T_r do not show the same tendency since none case reaches significance for DE (in fact only one value is found for IS in IV grade in *Bad + Poor*). Probably for younger children the adaptation under T_r or T_p is either masked by the causes of variability highlighted above (Sec. II A) or this trend stems from a combination of developmental skills which simply do not come into play until a certain maturity of the testers. In this second case and according to the present findings it seems reasonable to set an age reference close to 10 years. However, the final verification of this two hypotheses would require more specific tests and cannot be resolved within this data set.

Coming to DE for the activity noise, one finds not significant values for both IV and V grades, while the most telling results (though only significant at $\alpha = 0.1$) are for the positive correlations in III grade for *Bad* ($p = 0.052$, $\rho = 0.545$) and *Bad + Poor* ($p = 0.095$, $\rho = 0.315$) strata. But the analysis for IV and III grades becomes very interesting when RT data are considered (Table VII). In particular it is seen that for both grades A has significant decreasing trends with remarkable effect sizes, which are not found neither for other noises nor for the V grade under the same noise. The respective IS trends fail significance and thus for A only a "partial" adaptation involving RT can be depicted, and this is effective for younger children. Unfortunately the present statistics is not robust enough to establish a clear ranking between III and IV grade, but some indications from literature do help in formulating a possible explanation. In fact, as reported in the review,[5] inside primary classrooms the expected level of internal noise, which has the closest similarity with the present activity noise, increases with the decrease of the grade. Younger pupils are thus more exposed to this type of noise and, thanks to a known capability of gating a familiar noise disturbances,[29] they profit from a relative advantage during the stage of information processing. Nonetheless, it is to be remarked that this benefit is not sufficient to grant a better lesson-average performance to younger pupils compared to older ones, since the increase of proficiency with age has a stronger opposite effect on the average performance (see Sec. VI A).

V. OUTLINE OF THE ROLE OF SPEECH INTELLIGIBILITY AND RESPONSE TIME

In order to further explain the mechanisms behind the deterioration of performance during the lesson, the analysis was focused on the four significant cases highlighted by DE for the *Good + Fair* stratum for V (all three noises) and for IV (only T_r). To this aim the percentage losses with respect to the initial values were calculated from best fit regression lines in Figs. 2(a), 2(b), and 2(c) which are (DE), (IS), and (RT), respectively. The results are reported in Table VIII. It is to be recalled that a positive DE loss coincides with the previous decrease of listening efficiency, and the same concept is applied to IS. On the contrary a positive RT loss witnesses a quicker and thus easier reception process. It is also useful to recall that, as outlined in Ref. 8, when low-context test material is used,[30] the response time is more linked to the "top down" part of the speech understanding (i.e., filling the information gaps in the disrupted message for instance by using the language prior knowledge), whereas IS are more sensitive to the sensory cues (i.e., "bottom up" part of the cognitive process, that is building the message form phonemic and acoustic components). These basic criteria allow to identify the interplay of cognitive resources that build up pupils' response deterioration caused by noises during the lesson period.

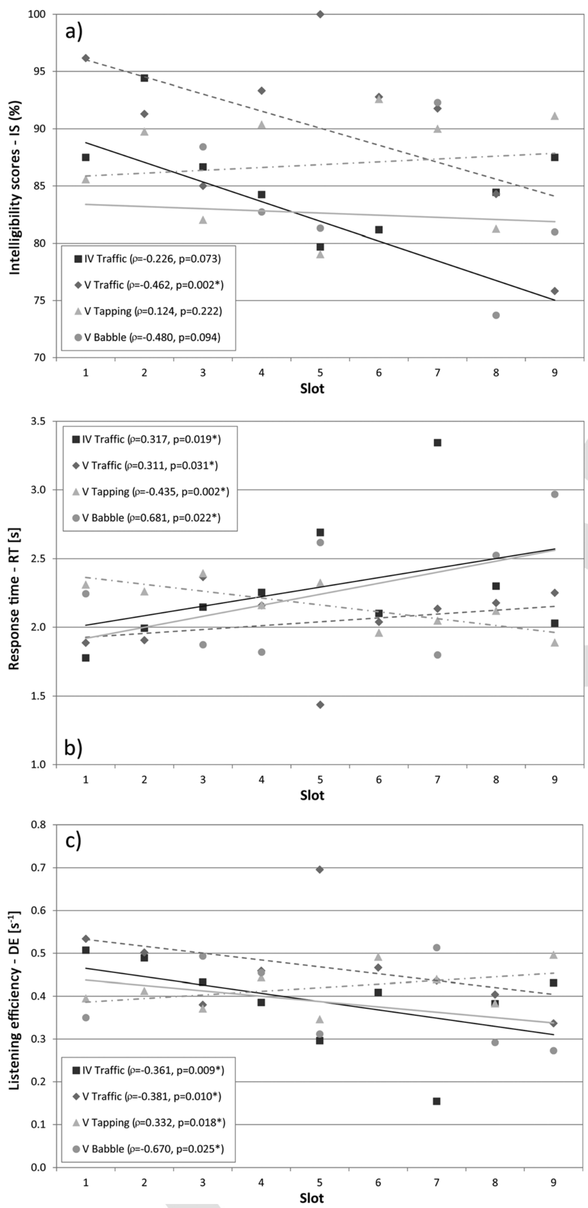

FIG. 2. Plots of the regressions obtained for the cases outlined in Table VIII for V and IV grades in the *Good + Fair* stratum. In (a) the listening efficiency DE, in (b) the speech intelligibility scores IS and in (c) the response time RT. Details on modification of data for presentation as in Fig. 1. Spearman's correlation ρ coefficients and the respective p values are included. An asterisk indicates significance at the $\alpha = 0.05$ level.

By considering in Table VIII the values of DE for V under T_r and A, one can see a loss for both, which is 23.1% for T_r and 17.2% for A, but the decrease in performance is realized in a peculiar way for each of the two noises. First, for T_r one finds a decrease of IS whilst for A the intelligibility scores are almost constant. The practical meaning of this finding can be understood by considering the ranges provided in Ref. 24, that match intelligibility scores with the rating of communication quality. In case of T_r, which starts at IS equal to 95% and then drops by 9.5% of this value, one would shift from a rating of *Good* to a *Fair* one during the lesson, whereas for A one would keep the *Fair* rating due to an almost constant IS behavior. In synthesis the continuous exposition to T_r noise stresses pupils' resources so that accuracy in resolving the energetic masking deteriorates during the lesson, while accuracy of reception under A is worse from the beginning but is only faintly affected by a continuous exposition.

Second, as regards RT, the initial values of both noises are surely above the reference ones for comfort. In fact the reference values, measured in anechoic conditions with SNR $\geq$ 30 dB, are (1.6 ± 0.3) s for V grade and (1.8 ± 0.3) s for IV grade.[20] During the lesson period T_r, starting from a lower point, gains 13.8% RT while A increases by 20.6%. This means that under both noises deciphering the available information during the lesson becomes more effortful, but pupils are capable to keep a satisfactory accuracy only in A, whereas the increased effort in T_r is not sufficient to avoid a drift toward lower IS.

Furthermore, it is interesting to note that under T_r also IV grade behaves similarly to grade V, even though the trend of IS loss is only significant at $\alpha = 0.1$. Keeping this limit in mind one can see that IS loss is 11.2% in IV compared to 9.5% in V, and that the RT loss is -17.6% in IV compared to -13.8% in V. Both changes seem amplified in IV compared to V and it can be hypothesized that these results stem from the peculiar group proficiency. More specifically this finding suggests that younger pupils generally achieve an inferior average performance (see Sec. VI A) due to a greater run-time deterioration of the basic speech reception mechanisms. Unfortunately this idea, which deserves future investigation, cannot be confirmed by the III graders because all of the respective trends fail significance due to the more scattered nature of the data.

Last, T_p for V grade in Table VII shows some 27% improvement in the listening efficiency during the lesson. The respective IS trend is not significant ($p = 0.222$), which keeps intelligibility in the *Fair* rating interval (80% < IS < 93%). RT decreases greatly and is thus fully responsible for the DE trend. So, contrary to A, under T_p the effort required to maintain accuracy is released during the lesson. This is a clear indication that such type of noise is more easily filtered by the V grade pupils or, in other words, that they appear resilient to T_p.

TABLE VIII. Percentage of loss of IS, RT and DE for T_r, T_p and A estimated from the best regression lines with respect to the initial value (reported in square brackets). Data refer to the subset of four significant cases of DE for V and IV in the *Good* + *Fair* stratum taken from Table IV. In two IS cases an asterisk (*) denotes significance at $\alpha = 0.1$. The former is V grade for A ($p = 0.094$, $\rho = -0.48$) and the latter is IV grade for T_r ($p = 0.073$, $\rho = -0.226$).

	Traffic (T_r)			Tapping (T_p)			Babble (A)		
Good + Fair	%IS loss	%RT loss	%DE loss	%IS loss	%RT loss	%DE loss	%IS loss	%RT loss	%DE loss
V	9.5 [95]	−13.8 [1.93]	23.1 [0.52]	−4.4 [85.5][a]	19.1 [2.38]	−27 [0.37]	−2.5 [80.3]*	−20.6 [2.07]	17.2 [0.41]
IV	11.2 [90]*	−17.6 [2.02]	24.3 [0.46]	—	—	—	—	—	—

[a]This trend is not significant ($p = 0.222$) and was reported for completeness. In the text IS is assumed constant across slots in this case.

VI. DISCUSSION

A. Merging with time-averaged listening efficiency results

The present findings on the DE run-time trends can be merged with the lesson-averaged DE results to fully describe the noise intrusion process. In Tables IX and X, the lesson-averaged results are resumed for the grades and for the noises, respectively. They are presented in the form of inequalities that describe the rank orders obtained by means of a dedicated statistical analysis comparing the distributions of the lesson-averaged listening efficiencies.[20] Both for noises and grades one can see that the better strata contribute in defining a clear ordering, whereas this does not happen in the worse acoustical conditions. Concerning the grades in Table IX, it is verified that the listening efficiency of V is better than both IV and III (column "*Fair* + *Good*") and just for noise T_r one finds equality between the last two. As detailed in Ref. 20, where a full discussion can be found together with the analysis of the separated IS and RT behaviors, the inequalities in Table IX can be justified with the improvement of the children skills stemming from the developmental process which is typical of the age range under investigation.

As regards the ranking of noises obtained with the lesson-averaged DE results, one can see in Table X that DE for A is worst for each grade and then T_p and T_r are always following in ascending order of performance. Then, in summary, the time-averaged lesson data show that, within the same objective STI ranges, a better performance is achieved in the classroom for T_r and lesser and lesser, respectively, for T_p and A.

In order to match the run-time and the time-averaged results presented above it has to be recalled that, for each noise and grade, one finds specific initial values of DE course at the start-up of the lesson. The time-average performance during T_r resulted the highest (that is $T_r > T_p > A$ last column of Table X), but run-time data showed that DE under T_r is subject to deterioration during the lesson (in particular IV–V grades, *Good* + *Fair*). Also in A (V grade) one finds a run-time decrease, but in this case the resulting average values are lowest. A more unclear view is depicted for T_p since results with an inverse trend are achieved by run-time analysis, whereas, as recalled above, the average behavior was validated as intermediate between A and T_r. Besides the type of noise, the trends are mediated also by the age and thus by the skill of the pupils. In particular, while T_r causes a drop of performance both in V and IV, not significant *p* results are achieved in the better conditions for III grade. Moreover under better acoustics the effect of A and T_p is significant for V grade only, but it is not for IV and III.

One can argue that the proficiency of pupils is manifested during the lesson as a peculiar performance reaction to the noise stimulus, and not only as higher or lower DE values when averaged during the whole lesson as the previous experiments showed. In particular older pupils seem capable of managing the noise intrusion more effectively at the start-up of the lesson period, since only due to a continuous exposition their performance deteriorates. As a result, their respective averaged performance is still better for each noise (Table IX, column *Fair* + *Good*), since younger students do not show a similar behavior.

B. Effects on pupils

The previous analysis demonstrates that the allocation of the available cognitive resources that older pupils can implement to cope for the noise do vary during a lesson period, is adapted to the input conditions and is not equally effective for the three types of noise considered in the above experiments. In particular, as explained in Sec. V when considering Table VIII, V grade pupils are able to substantially keep a less-than-optimal accuracy as described by intelligibility scores but at the expense of an increased effort as

AQ4

TABLE IX. (From Ref. 20): Listening efficiency (DE); results of the ordering of the lesson-averaged data for the noises. The abbreviations T_r, T_p and A represent the respective statistical distributions and the subscript refers to the grade. The inequalities are referred to the statistical distributions of listening efficiency (DE) and are obtained by stochastic ordering procedure (Ref. 26). Full ordering is achieved with darker grey background and partial ordering with a lighter one.

	Strata				
Noise type	Bad	Poor	Fair	Good	Fair + Good
T_r	$Tr_{III} = Tr_{IV} = Tr_V$	$Tr_{III} = Tr_{IV} = Tr_V$	$\mathbf{Tr_{III} = Tr_{IV} < Tr_V}$	$Tr_{III} = Tr_{IV} = Tr_V$	$\mathbf{Tr_{III} = Tr_{IV} < Tr_V}$
T_p	$Tp_{III} = Tp_{IV} = Tp_V$	$\mathbf{Tp_{III} = Tp_{IV} < T_{pV}}$	$\mathbf{Tp_{III} < Tp_{IV} = Tp_V}$	$\mathbf{\underline{Tp_{III}} < \underline{Tp_{IV}} < \underline{Tp_V}}$	$\mathbf{\underline{Tp_{III}} < \underline{Tp_{IV}} < \underline{Tp_V}}$
A	$A_{III} = A_{IV} = A_V$	$\mathbf{A_{III} = A_{IV}} < \mathbf{A_V}$	$\mathbf{A_{III}} < \mathbf{A_{IV} = A_V}$	$\mathbf{A_{III} = A_{IV}} < \mathbf{A_V}$	$\mathbf{\underline{A_{III}} < \underline{A_{IV}} < \underline{A_V}}$

ID: **aipepub3b2server** Time: 20:25 I Path: D:/AIP/Support/XML_Signal_Tmp/AI-JAS#150615

AQ5 TABLE X. (Adapted from Ref. 20) Listening efficiency (DE); results of the ordering of the lesson-averaged data for the classes. Abbreviations and meaning of symbols as in Table IX.

Grade	Strata				
	Bad	Poor	Fair	Good	Fair + Good
III	$A = T_p = T_r$	$A = T_p = T_r$	**$A < T_p = T_r$**	**$A = T_p < T_r$**	**$\underline{A} < \underline{T_p} < \underline{T_r}$**
IV	$A = T_p = T_r$	$A = T_p = T_r$	**$A < T_p$**	**$A = T_p < T_r$**	**$\underline{A} < \underline{T_p} < \underline{T_r}$**
V	**$T_r < T_p = A$**	$A = T_p = T_r$	**$A = T_p < T_r$**	$A = T_p = T_r$	**$\underline{A} < \underline{T_p} < \underline{T_r}$**

response time indicates, while T_r is detrimental for both IS and RT. An increase in the RT can be interpreted as a symptom of "fatigue" since the same task requires longer time to be completed. Toward the end of the period a cumulative effect can be expected, so that exertion will impede to follow the lesson and cause the probable giving up of the students. If this deteriorating mechanism becomes customary in the school experience of children it can severely expose them to the occurrence of "learned helplessness" which is well documented in the literature.[29]

Moreover, the present findings are coherent with previous studies which used physiological measures, such as cortisol levels, to evaluate fatigue in school children[6] and which found that classroom noise level is related to stress reactions among children, such as fatigue and headache and a reduced diurnal cortisol variability.[31]

While there is a consensus on the tiring effect of noise on children performance (even with exceptions due to possible arousal effects in some cases) the interpretation of the adaptation to acute noise (that is not chronic) is not as clear. Few studies outline adaptation after a short exposition, whereas others affirm the absence of adaptation.[32] The explanation lays in the presence or not of short breaks with relatively calm conditions which are fit to clear the adaptation process. In the present experiments breaks were provided only after 45 to 55 min, so that the rather long session proposed realized conditions compatible with the appearance of adaptation.

Under worse acoustical conditions the adaptation, when applicable, cannot be considered as a strategy to prevent the pupils from the adverse effects of noise on performance in word recognition. In fact, with reference to Table V for T_r in grade V, *Bad* stratum, one finds that in this case, which is significant for all of the three quantities, the respective intelligibility scores increase from 53% to 66% thus shifting from a *Bad* to a *Poor* rating,[24] which is obviously not acceptable as a target for speech intelligibility in the classrooms. A partial adaptation was also found for III and IV grade pupils under worse activity noise since response time had a decreasing trend. This finding was traced back to the more recent exposition of younger classes to a louder version of a similar type of noise. Although the finding is intriguing, the intelligibility scores were not affected and listening efficiency was only faintly touched. Thus, the effective benefit for the pupils is difficult to estimate.

In a more general perspective the issues highlighted so far link acoustics, lesson organization and the strategies to control the adverse effects of noises on students. For instance, in better but still not optimal conditions the management of short breaks during the lesson may compensate somehow the expected decrease in performance, thus mitigating for the pupils the effect of noise and reverberation. This will be most effective for higher external traffic noise and for internal activity noise levels especially for older groups. The former case applies for classrooms directly exposed to road noise in dense urban areas, whereas the latter is typical of lesson styles involving group work, which is more and more valorized besides frontal lesson. Fortunately, tapping noise from the upper floor, whose control is hardly possible by the teacher of the disturbed class, seems not as critical as the other two. On the other hand, an effective management of classroom noise by the teachers requires both awareness of the potential differential impact of noises across tasks and a specific training to modulate the effects of noise; in one specific research[33] both aspects showed large room for improvement.

VII. CONCLUSIONS

Comfortable listening in the educational premise is mandatory for learning, and this prompted to investigate how performance, qualified here by means of the number of words correctly understood within 1 s, is developed under noise conditions during a lesson both on a run-time and on a time-average basis. Similar or equal time-average values of listening efficiency are found to be realized in peculiar ways depending on the match of run-time speech intelligibility scores and response times. The joint use of the two quantities is able to depict the impact of noise on the masking of the signal and on the cognitive resources involved in the deciphering process. In summary the main findings of the work can be outlined as follow:

- Better listening conditions: probably due to the nature of the data set, "fatigue" was only verified for older pupils in some cases and, based on the number of congruent significant results, also a dependence on the type of noise was observed. In particular T_r and A are consistent in the decrease of efficiency, whereas T_p has a prevalent opposite trend (see Tables V and VIII, V grade, *Good + Fair*) which is driven by a reduced RT and seems to indicate that they are more resilient to this type of noise.
- Worse listening conditions: in this cases "adaptation" may occur, that is an increase of listening efficiency during the lesson period, and this was strictly reported for T_r only in the worst strata for grade V, whilst similar cases in IV and III grades failed the significance testing. A partial "adaptation" concerning A noise in grades III and IV was reported and this was considered as an indication that a

more recent past exposure to a similar noise can have a measurable effect on children's effort in the speech reception process under noise.

- Aside the peculiar case of A where a partial trend was depicted, younger pupils (III grade) failed to show significant trends in listening efficiency in either better or worse conditions. The reason for this occurrence needs more specific evaluation; altogether the data presented in Sec. V for T_r noise between V and IV seem to provide a description of the way the proficiency of pupils is developed with the age. In particular an enhancement of IS loss (less accuracy) and RT increase (more effort) in IV grade with respect to V grade was reported in Table VIII for T_r. Thus one can argue that an improvement of the resiliency to noise during the course of the lesson should be considered as part of the developmental proficiency in speech reception under adverse conditions.

The subjective responses of pupils collected by accuracy and latency measures are the results of acoustically complex and interlinked phenomena, whose simultaneous control is problematic in the working classroom. In fact temporal fine structure, frequency span, intensity and spatial attributes of noise are all involved at various degree in the noise intrusion effect, and also the attention-capture potential of noise is especially important for children. Thus, since the standardized tools such as STI are restricted to the energetic masking acted by noise on speech,[34] they are not entirely adequate to provide a complete description of the phenomenon, and further or alterative information is necessary. Different and sophisticated methods to predict speech intelligibility have been developed, also for resolving some of the limits of STI under critical conditions (see for instance Refs. 35 and 36) and also a much simplified approach was depicted for classrooms.[37] Nonetheless, no method is currently available to predict in a simple and reliable way the effort put in the reception process. An objective design criteria correlating not only with accuracy (i.e., items correctly judged) but with the subjective performance under various noise conditions would be an extremely valuable tool in the room acoustical design process when optimizing spaces for learning. In fact it is believed that, in a perspective, future regulations shall consider not only energetic masking or reverberation time limits but shall specify which features of noise and to what extent they have to be controlled in order to warrant a given level of performance in the classroom.

[1]B. M. Shield and J. E. Dockrell, "The effects of noise on children at school: A review," Build. Acoust. **10**(2), 97–116 (2003).
[2]K. Anderson, "The problem of classroom acoustics: The typical classroom soundscape is a barrier to learning," Sem. Hearing **25**(2), 117–129 (2004).
[3]B. M. Shield and J. E. Dockrell, "The effects of environmental and classroom noise on the academic attainments of primary school children," J. Acoust. Soc. Am. **123**(1), 133–144 (2008).
[4]M. Klatte, K. Bergström, and T. Lachmann, "Does noise affect learning? A short review on noise effects on cognitive performance in children," Front. Psychol. **4**, 578 (2013).
[5]M. Picard and J. S. Bradley, "Revisiting speech interference in classrooms," Int. J. Audiol. **40**(5), 221–244 (2001).
[6]C. B. Hicks and A. M. Tharpe, "Listening effort and fatigue in school-age children with and without hearing loss," J. Speech Language Hearing Res. **45**, 573–584 (2002).
[7]C. S. Howard, K. J. Munro, and C. J. Plack, "Listening effort at signal-to-noise ratios that are typical of school classrooms," Int. J. Audiol. **49**(12), 928–932 (2010).
[8]N. Prodi, C. Visentin, and A. Farnetani, "Intelligibility, listening difficulty and listening efficiency in auralized classroom," J. Acoust. Soc. Am. **128**(1), 172–181 (2010).
[9]M. Klatte, J. Hellbrück, J. Seidel, and P. Leistner, "Effects of classroom acoustics on performance and well-being in elementary school children: A field study," Environ. Behav. **42**(5), 659–692 (2010).
[10]L. Valente, H. M. Plevinsky, J. M. Franco, E. C. Heinrichs-Graham, and D. E. Lewis, "Experimental investigation of the effects of the acoustical conditions in a simulated classroom on speech recognition and learning in children," J. Acoust. Soc. Am. **131**(1), 232–246 (2012).
[11]S. Hygge, "Classroom acoustics–Speech intelligibility, memory and learning," in *Proceedings of Forum Acusticum*, Krakow, Poland (September 7–12, 2014).
[12]R. Ljung and A. Kjellberg, "Long reverberation time decrease recall of spoken information," Build. Acoust. **16**(4), 301–311 (2009).
[13]R. Ljung, P. Sörquist, A. Kjelberg, and A.-M. Green, "Poor listening conditions impair memory for intelligible lectures: Implications for acoustic classroom standards," Build. Acoust. **16**(3), 257–265 (2009).
[14]R. McGarrigle, K. J. Munro, P. Dawes, A. J. Stewart, D. R. Moore, J. G. Barry, and S. Amitay, "Listening effort and fatigue: What exactly are we measuring?," Int. J. Audiol. **53**(7), 433–445 (2014).
[15]C. Pals, A. Sarampalis, H. van Rijn, and D. Başkent, "Validation of a simple response-time measure of listening effort," J. Acoust. Soc. Am. **138**, EL187 (2015).
[16]IEC 60268-16:2011: *Sound System Equipment—Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index* (International Electrotechnical Commission, Geneva, Switzerland).
[17]J. Rennies, H. Schepker, I. Holube, and B. Kollmeier, "Listening effort and speech intelligibility in listening situations affected by noise and reverberation," J. Acoust. Soc. Am. **136**(5), 2642–2653 (2014).
[18]H. Sato, J. S. Bradley, and M. Morimoto, "Using listening difficulty ratings of conditions for speech communication in rooms," J. Acoust. Soc. Am. **117**(3), 1157–1167 (2005).
[19]M. Klatte, T. Lachmann, and M. Meis, "Effects of noise and reverberation on speech perception and listening comprehension of children and adults in a classroom-like setting," Noise Health **12**, 270–282 (2010).
[20]N. Prodi, C. Visentin, and A. Feletti, "On the perception of speech in primary school classrooms: Ranking of noise interference and of age influence," J. Acoust. Soc. Am. **133**(1), 255–268 (2013).
[21]P. Bonaventura, A. Paoloni, F. Canavesio, and P. Usai, "Realizzazione di un test diagnostico di intelligibilità per la lingua italiana (Development of a diagnostic intelligibility test in the Italian language)," Internal Technical Rep. No. 3C1286, Fondazione Ugo Bordoni, Rome, 1986.
[22]F. Cutugno, S. Prosser, and M. Turrini, *Audiometria Vocale (Vocal Audiometry)* (GN ReSound, Padua, Italy, 2000), Vol **1**.
[23]W. A. Dreschler, H. Verschuure, C. Ludvigsen, and S. Westermann, "ICRA (International Collegium for Rehabilitative Audiology) noises: Artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment," Audiology **40**, 148–157 (2001).
[24]ISO9921: *Ergonomics–Assessment of Speech Communication* (International Organization for Standardization Geneva, Switzerland, 2003).
[25]N. Prodi, C. Visentin, and C. Bellettini, "Listening efficiency testing," in *Proceedings of 45th Audio Engineering Society Conference*, Helsinki, Finland (March 1–4, 2012).
[26]S. Bonnini, N. Prodi, L. Salmaso, and C. Visentin, "Permutation approaches for stochastic ordering," Commun. Stat. **43**(10–12), 2227–2235 (2014).
[27]N. Prodi, C. Visentin, and A. Feletti, "Fatigue and adaptation in noisy primary schools," in *Proceedings of EURONOISE2012*, Prague, Czech Republic (June 10–13, 2012).
[28]J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. (Lawrence Erlbaum Associates, New York, 1988), Chap. 3.
[29]G. W. Evans and S. J. Lepore, "Nonauditory effects of noise on children: A critical review," Children's Environ. **10**(1), 31–51 (1993).
[30]D. N. Kalikow, K. N. Stevens, and L. L. Elliott, "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," J. Acoust. Soc. Am. **61**, 1337–1351 (1977).
[31]R. Wålinder, K. Gunnarsson, R. Runeson, and G. Smedje, "Physiological and psychological stress reactions to classroom noise," Scand. J. Work Environ. Health **33**(4), 260–266 (2007).
[32]S. P. Banbury, W. J. Macken, S. Tremblay, and D. M. Jones, "Auditory distraction and short-term memory: Phenomena and practical implications," Human Factors **43**(1), 12–29 (2001).

[33]J. E. Dockrell, B. M. Shield, and K. Rigby, "Acoustic guidelines and teacher strategies for optimizing learning conditions in classrooms for children with hearing problems," in *Proceedings of the 1st Conference ACCESS: Achieving Clear Communication Employing Sound Solutions* (Phonak AG, Chicago IL, 2003), pp. 217–227.

[34]P. Larm and V. Hongisto, "Experimental comparison between speech transmission index, rapid speech transmission index, and speech intelligibility index," J. Acoust. Soc. Am. **119**(2), 1106–1117 (2006).

[35]M. Lavandier, S. Jelfs, J. F. Culling, A. J. Watkins, A. P. Raimond, and S J. Makin, "Binaural prediction of speech intelligibility in reverberant rooms with multiple noise sources," J. Acoust. Soc. Am. **131**(1), 218–231 (2012).

[36]S. Jorgensen, S. D. Evert, and T. Dau, "A multi-resolution envelope-power based model for speech intelligibility," J. Acoust. Soc. Am. **134**(1), 436–446 (2013).

[37]C. Visentin and N. Prodi, "A comparative analysis between objective and psychoacoustics indicators describing speech reception in primary school classrooms," in *Proceedings of INTERNOISE2013*, Innsbruck, Austria (September 15–18, 2013).