

Machine learning per la separazione e la misura di sorgenti sonore coesistenti in spazi chiusi

Domenico De Salvio¹

Dipartimento di Ingegneria Industriale,
Università di Bologna,
Viale del Risorgimento, 2, 40136 Bologna
domenico.desalvio2@unibo.it

Ricevuto: 28/2/2024

Accettato: 20/5/2024

DOI: 10.3280/ria1-2024oa17373

ISSN: 2385-2615

La crescente attenzione verso i monitoraggi a lungo termine, e alle grandi quantità di dati che riescono a produrre, spiana la strada all'utilizzo di tecniche di machine learning. L'utilizzo di algoritmi sofisticati, principalmente basati su tecniche statistiche, permette di ampliare le capacità di analisi dei tecnici acustici in contesti complessi. Il presente lavoro vuole proporre un metodo basato su tecniche di clustering per separare, identificare e misurare diverse sorgenti sonore coesistenti in scenari reali monitorati tramite un fonometro. Sono presentati quattro casi studio in due contesti diversi (aule universitarie e uffici) in cui il metodo proposto è stato applicato.

In entrambi i contesti, i risultati hanno evidenziato come il clustering riesca a ricostruire in modo efficace le occorrenze misurate, restituendo un significato fisico ad ogni cluster attraverso match spettrali e parametri statistici e temporali. Le indagini effettuate nei casi studio presentati promuovono l'uso del metodo proposto come un possibile strumento per ricavare ulteriori informazioni in scenari complessi.

Parole chiave: machine learning, monitoraggi, fonometro, aule universitarie, uffici, misura, spazi chiusi

Machine learning for the separation and measurement of coexisting sound sources in enclosed spaces

The constant growth of the interest in long-term monitoring and the large amount of data it produces pave the way for machine learning techniques. Using sophisticated algorithms, usually based on statistics, broadens the capability of analysis in complex scenarios. The present work proposes a machine learning-based method to separate, identify, and measure different coexisting sound sources through sound level meters in real-world scenarios. Four case studies show the application of the proposed method in two different contexts (university lecture halls and offices). In both contexts, results show the reliability of the clustering technique. It efficiently reconstructs the occurrences measured and gives back a physical meaning to clusters through spectral matching and statistical and temporal parameters. The investigation presented here fosters the use of the proposed method as a possible tool to mine information in complex scenarios.

Keywords: machine learning, monitoring, sound level meter, university lecture hall, offices, measurement, enclosed spaces

1 | Introduzione

Al giorno d'oggi la prassi tecnica comune fra i tecnici acustici si basa sulla descrizione dei contesti sonori attraverso i livelli di pressione sonora (SPL). Questi possono essere raccolti attraverso un fonometro e descrivono il contesto principalmente attraverso metriche energetiche.

Il parametro più comunemente utilizzato è il livello equivalente (L_{eq}), ovvero il valore quadratico medio valutato su un determinato intervallo di tempo. Tuttavia, essendo una metrica energetica, i livelli più alti influiscono significativamente sui risultati. Il L_{eq} viene principalmente utilizzato per valutare l'espo-

sizione sonora a lungo termine durante periodi rappresentativi di misurazione. Per eliminare l'influenza della durata del tempo di misurazione si usa il Sound Exposure Level (SEL) [1]. In questo caso, si "schiaccia" l'informazione misurata su un tempo di riferimento pari a 1 s. Tuttavia, è possibile utilizzare i dati fonometrici anche da un punto di vista statistico, oltre che energetico. Infatti, analisi più dettagliate vengono effettuate attraverso livelli di pressione statistici. I percentili acustici (L_N) sono definiti come gli SPL superati per l' $N\%$ del tempo di misura, dove N è la percentuale rispettiva da considerare. Ad esempio, L_{90} rappresenta l'SPL superato per il 90% del tempo di misura. I livelli percentili vengono comunemente usati per descrivere i contributi di singole sorgenti. Ad esempio, L'_{90} viene comunemente associato al livello di rumore di fondo. Tuttavia, questo approccio si basa su assunzioni temporali difficilmente dimostrabili.

¹ Vincitore ex-aequo del premio "Gino G. Sacerdote" edizione 2024.

La prassi attuale non permette di misurare in un contesto reale i diversi contributi delle diverse tipologie di sorgenti esistenti. Questa possibilità permetterebbe analisi più accurate per l'analisi e la progettazione acustica in molti casi.

Diversi lavori hanno indagato la possibilità di separare le sorgenti sonore attraverso monitoraggi a lungo termine [2–13]. Siccome da questi è possibile ricavare una grande quantità di dati, l'approccio statistico si è rivelato più efficace di quello energetico. L'utilizzo di SPL come variabili casuali porta a descrivere i contesti di misura attraverso i suoi SPL più probabili. Ciò non implica che i due approcci vadano in contrasto fra di loro. Misurano e descrivono il contesto sonoro secondo caratteristiche diverse.

La possibilità di usufruire di larghe quantità di dati ha permesso negli ultimi anni l'esplorazione sempre più frequente di nuove tecniche di analisi dati derivanti dal campo del machine learning [14]. Applicazioni di tecniche di machine e deep learning in acustica si trovano nei più svariati campi, dalle analisi geotecniche alle applicazioni in ambienti riverberanti [15–17]. Tuttavia, queste tecniche richiedono un alto costo di calcolo e dunque strumentazioni molto performanti poiché utilizzano input ricchi di informazioni come ad esempio spettrogrammi, MFCC, o direttamente registrazioni audio [18]. L'utilizzo dei soli SPL come input permette invece di ottenere risultati con costi di calcolo inferiori. Esempi di applicazione di approcci statistici basati solo su SPL li troviamo in contesti come aule scolastiche, uffici, ospedali ed eventi sportivi.

Il presente lavoro vuole proporre un approccio statistico dell'uso di monitoraggi fonometrici attraverso un metodo rigoroso e validato sia teoricamente che sperimentalmente.

2 | Accenni teorici

2.1 | Clustering

Il metodo proposto è basato sulla collezione di SPL ottenuta tramite monitoraggi fonometrici a medio e lungo termine. Ciò significa che il database di partenza per l'analisi da svolgere è formato da una sola lunga serie di numeri. Non ci sono dunque informazioni riguardo i numeri collezionati. Nell'ambito dell'analisi dati, si dice che i dati non sono "etichettati". Dunque, il primo passo che si può compiere è svolgere un'analisi di clustering.

Il clustering è una tecnica di analisi dati che impiega degli algoritmi per ricercare pattern all'interno dei dati [19]. L'obiettivo è suddividere i dati che si hanno a disposizione in diversi gruppi, chiamati *cluster*, in base alla loro somiglianza. L'ottimizzazione sarà poi ottenuta tramite processi iterativi che si concentreranno sulla coesione dei gruppi ricavati. Il miglior risultato possibile considerando i parametri di somiglianza e coesione rappresenterà il risultato finale. L'obiettivo finale del clustering è ricavare caratteristiche utili all'interno dei gruppi che saranno chiamate *features*. Questo processo è anche comunemente chiamato, nell'ambito del machine learning, apprendimento non supervisionato (*unsupervised*) [14,20].

Ogni algoritmo in grado di condurre un'analisi di clustering forma i cluster in base a criteri differenti. Nel presente lavoro si utilizzerà un algoritmo detto *model-based*, ciò significa che modella i dati da un processo generativo. Una volta scelto il processo generativo, si possono stimare i parametri del modello e ricavarne dunque le *features*.

2.2 | No free-lunch theorem e scelta dell'algoritmo

Il *no free-lunch theorem* (NFL) sostiene che, nello spazio di tutti i possibili problemi, ogni algoritmo di ottimizzazione produrrà in media risultati simili [21]. Una delle implicazioni di questo teorema è che non esiste un algoritmo migliore di un altro. Tuttavia, la conoscenza del contesto del problema da ottimizzare può influenzare la scelta dell'algoritmo.

Nel caso specifico del presente lavoro, il dato è rappresentato da una serie di short-time L_{eq} misurati da un fonometro sulla base del suo intervallo e costante di acquisizione. Ad esempio, un'acquisizione fonometrica con costante Fast sarà definita come:

$$SPL = 10 \log \left[\frac{(1/\tau_F) \int_{-\infty}^t p^2(\xi) e^{-(t-\xi)/\tau_F} d\xi}{p_0^2} \right] \text{ (dB)} \quad (1)$$

Dove τ_F rappresenta la costante Fast, ξ la variabile fittizia di integrazione, p la pressione istantanea misurata, p_0 la pressione di riferimento. Al netto della costante e dell'intervallo di acquisizione, ciò che è importante notare è che il dato è rappresentato da una quantità al quadrato all'interno di un logaritmo. Questo, nell'ambito dell'analisi dati, rappresenta ciò che viene definita una *log-transformation* ed è una tecnica comune per trasformare una distribuzione qualunque di dati in una con andamento normale [22]. Ciò significa che, a livello teorico, gli SPL prodotti da una singola sorgente, monitorata per un intervallo di tempo relativamente lungo, formeranno una curva di distribuzione gaussiana.

2.3 | Gaussian Mixture Model

Partendo dalle considerazioni fatte nella sezione precedente, l'algoritmo scelto per svolgere l'analisi di clustering su SPL sarà il Gaussian Mixture Model (GMM) [23]. Questo è un modello probabilistico che ricostruisce la struttura originale dei dati iniziali attraverso una combinazione lineare di distribuzioni gaussiane. L'ottimizzazione avviene tramite un processo iterativo che aggiorna gli *iperparametri*, cioè le medie, le covarianze, e i pesi di ogni componente della mistura. Dato un set di N osservazioni indipendenti $x = \{x_1, \dots, x_N\}$, la mistura gaussiana è definita come:

$$f(x) \approx \sum_{k=1}^K \pi_k f_k(x_i, \mu_k, \sigma_k^2) \quad (1)$$

dove π_k rappresenta i pesi di ogni componente, definiti come:

$$0 \leq \pi_k \leq 1 \quad (k = 1, \dots, K) \quad (2)$$

e

$$\sum_{k=1}^K \pi_k = 1. \quad (3)$$

L'Eq. (1) è soddisfatta massimizzando la funzione di verosimiglianza, definita come:

$$\mathcal{L}(x) = \prod_{i=1}^n \sum_{k=1}^K \pi_k f_k(x_i) = \prod_{i=1}^n \sum_{k=1}^K \pi_k \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(x_i-\mu_k)^2}{2\sigma_k^2}}. \quad (4)$$

La massimizzazione della funzione di verosimiglianza è un metodo utilizzato in statistica per trovare i valori dei parametri che possono aver prodotto con la maggiore probabilità i dati osservati. Nel presente lavoro questo metodo è applicato alle distribuzioni normali, come esplicitato dall'ultimo termine indicato nell'Eq. (4) [24].

Nel presente lavoro, la massimizzazione viene fatta attraverso l'algoritmo Expectation-Maximization (EM) [25]. La Fig. 1 mostra un esempio di analisi GMM effettuata nel contesto del presente lavoro. La linea blu rappresenta la distribuzione delle occorrenze degli SPL acquisiti, mentre le 3 linee arancione, gialla e verde rappresentano 3 differenti cluster da associare a 3 differenti tipologie di sorgenti.

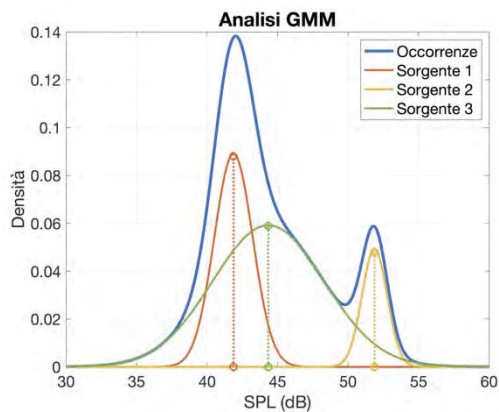


Fig. 1 – Esempio di ricerca dei cluster tramite GMM
Example of cluster analysis via the GMM

2.4 | Candidate models e numero ottimale di cluster

In molti algoritmi, tra cui la GMM, il numero di cluster da ricavare bisogna definirlo prima di avviare il processo iterativo. Questa rappresenta una delle tematiche più discusse nell'ambito delle analisi clustering [19]. La ricerca del numero ottimale di cluster si svolge su due step:

1. Creazione dei *candidate models*: si utilizza la GMM sui dati di partenza impostando diversi numeri di cluster k da ricercare. Ogni GMM(k) rappresenterà un candidato.
2. Si utilizza una metrica che valuta separazione e coesione dei vari cluster creati per ogni candidato. Il modello che ottiene il risultato migliore rappresenterà il k ottimale all'interno dei dati.

Esistono diverse metriche per valutare i candidati e anche fra di esse, in accordo con il NFL, non ne esiste una che performerà meglio di altre. Una delle tecniche più comuni per ovviare a questo limite è utilizzare più metriche e confrontarne i risultati.

Nel presente lavoro le metriche di selezione dei candidati sono quattro: l'indice *Calinski-Harabasz*, l'indice *Davies-Bouldin*, il coefficiente di *Silhouette*, e il *Gap Statistic*. Ognuna di queste metriche valuta separazione e coesione dei vari modelli in modo diverso [26-29]. Il numero ottimale di cluster più frequente fra le metriche utilizzate corrisponderà alla scelta finale. Approfondimenti teorici delle singole metriche utilizzate sono rimandati alla letteratura di riferimento.

3 | Metodo

3.1 | Casi studio

Il metodo proposto è stato costruito generalizzando tecniche avanzate di misure fonometriche in diversi contesti [4,6-9,12]. In questo lavoro si vogliono presentare i risultati di differenti casi studio affrontati in due differenti contesti di seguito descritti brevemente.

In Fig. 2 sono rappresentati due esempi di analisi GMM secondo il metodo proposto in questo lavoro, una svolta in un'aula universitaria, l'altra svolta in un ufficio. Di seguito sono descritti alcuni dettagli dei casi studio qui presentati.

3.1.1 | Aule universitarie

Il primo caso studio è rappresentato da due aule universitarie sottoposte ad interventi di miglioramento acustico secondo UNI11532-2 [30]. Le due aule sono considerabili come "grandi" avendo un volume rispettivamente di 900 e 1000 m³, con una capienza massima di 200 e 250 studenti. Qui il metodo proposto è stato utilizzato per valutare le variazioni di *student activity* (SA) in 18 lezioni totali, 9 prima e 9 dopo gli interventi realizzati. La SA è definita come il chiacchiericcio fra gli studenti durante le lezioni ed è considerata in letteratura una metrica dell'attenzione degli studenti e strettamente connessa alle proprietà acustiche delle aule [3]. SA viene ricavata tramite GMM e separata rispetto a quello che è il parlato dell'insegnante (indicato come *speech level* SL). È importante far notare che, essendo entrambe due aule grandi, gli insegnanti utilizzano un impianto di amplificazione per parlare agli studenti.

3.1.2 | Uffici

Il contesto degli uffici è qui affrontato tramite due casi studio. Due uffici sono stati monitorati al fine di ricavare i contributi spettrali di ogni tipologia di sorgente, come il parlato, il traffico, e il rumore di impianti. Gli spazi oggetto di studio sono abbastanza grandi da ospitare rispettivamente 4 e 3 postazioni di lavoro. In entrambi i casi studio il lavoro svolto è di tipo collaborativo. Ciò significa che i lavoratori parlano fra di loro oltre che con gente esterna attraverso canali tematici.

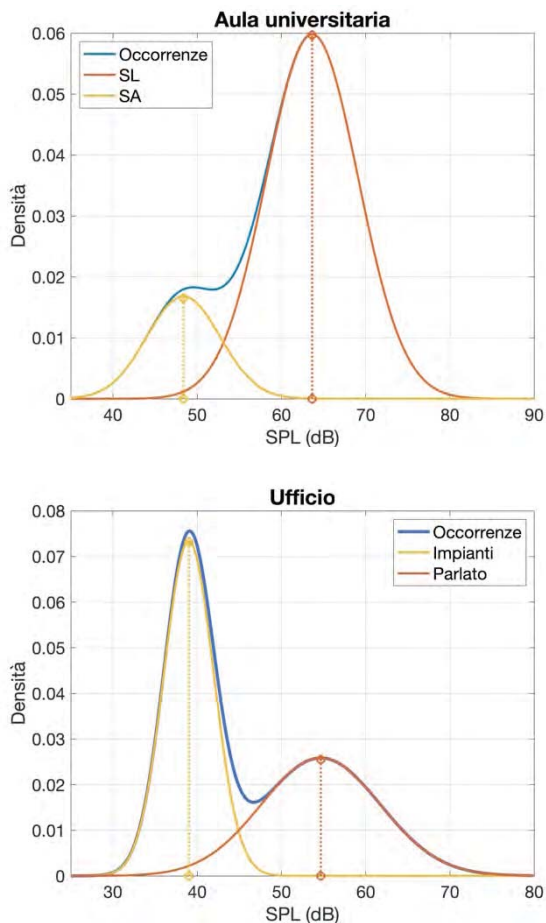


Fig. 2 – Esempi di analisi GMM in due contesti differenti: aula universitaria (in alto) e ufficio (in basso)
Example of GMM analysis in two different environments: university lecture hall (on the top) and office (on the bottom)

3.2 | Analisi dei cluster

Tutte le analisi sono basate sulla curva delle occorrenze dei monitoraggi effettuati. Questo permette di utilizzare il metodo proposto in base al contesto di applicazione e allo scopo della ricerca.

L'analisi può essere generalizzata e descritta in 5 step:

1. **Acquisizione dei dati.** Tutte le analisi presentate qui sono basate su monitoraggi fonometrici ottenuti con un intervallo di acquisizione di 100 ms in terzi e bande d'ottava. Ogni banda di frequenza, a prescindere dalla sua larghezza, influente nel metodo, rappresenterà una popolazione statistica. Ad esempio, tutti gli SPL acquisiti nella banda di ottava dei 125 Hz o nella banda di terzi di ottava dei 160 Hz rappresenteranno due differenti popolazioni statistiche da poter analizzare singolarmente.
2. **Preparazione dei candidate models.** Ogni singola popolazione statistica, i.e., ogni banda o livello globale del tempo di misura da voler analizzare, è elaborata più volte tramite GMM impostando diversi numeri di cluster k . Ad esempio, in Fig. 1, i dati sono stati suddivisi con $k=3$, in Fig. 2 con $k=2$. Questo passaggio è necessario perché

il numero di cluster k deve essere impostato prima di avviare il processo iterativo di GMM; dunque, non si conosce il numero ottimale di cluster all'interno delle varie popolazioni statistiche. A questo punto si ha una serie di popolazioni di dati suddivisi più volte con diversi k chiamati *candidate models*.

3. **Numero ottimale di cluster.** Fra i *candidate models* ce n'è uno che è quello ottimale. Il numero di cluster che avrà il modello ottimale rappresenterà il numero ottimale di cluster. Per conoscere questo numero, i.e., il numero di sorgenti sonore all'interno dei dati acquisiti, è necessario utilizzare le 4 diverse metriche citate in Sezione 2.4 per ogni gruppo di *candidate models* preparati nello step precedente. Poiché l'utilizzo di queste metriche è non supervisionato, è utile valutare i risultati ottenuti non soltanto tramite gli algoritmi utilizzati ma ponderando i risultati ottenuti con l'esperienza e la conoscenza del contesto. Per le metriche di valutazione, ognuna di esse indicherà dunque un numero ottimale di cluster. Il risultato finale sarà dato dal numero più frequente ottenuto; dunque, indicheranno il numero di sorgenti all'interno dei dati. Ad esempio, se si sta analizzando la popolazione statistica della banda di ottava dei 125 Hz, ogni metrica valuterà qual è la migliore fra le popolazioni dei 125 Hz elaborate con diversi k . Se due metriche indicheranno che il modello ottimale è quello ricostruito con $k=3$, una con $k=4$ e una con $k=2$, allora il modello ottimale sarà rappresentato da $k=3$.
 4. **Separazione delle sorgenti.** In questo step si separano i diversi cluster – cioè sorgenti – ottenuti per ogni banda di frequenza. Questo step colleziona le diverse medie e deviazioni standard di ogni cluster ottenuto per ogni popolazione statistica.
 5. **Identificazione delle sorgenti.** I cluster ottenuti sono rappresentati da curve gaussiane. Dunque, saranno definite univocamente da due parametri: la media e la deviazione standard. La media rappresenterà l'SPL della sorgente. La deviazione standard di ogni cluster è utilizzata per identificare il tipo di sorgente. Per esempio, prendendo a riferimento la distribuzione delle occorrenze graficata in basso della Fig. 2 si vedranno due tipi di cluster diversi. Le sorgenti di natura meccanica, avendo dei cicli ripetitivi nel tempo, emetteranno SPL in un range ristretto. Dunque, saranno costituite da curve gaussiane relativamente strette (curva gialla). Al contrario, il parlato di una persona, essendo un tipo di segnale molto vario nel tempo, dettato da pause e diverse intonazioni vocali, costituirà una curva gaussiana più larga (curva arancione). In questo lavoro, sulla base dei lavori precedentemente citati, si utilizzerà preliminarmente una soglia uguale a 5 dB per indicare una sorgente di tipo antropico (> 5 dB) da una di tipo meccanico o di traffico (≤ 5 dB).
- È importante sottolineare come questo metodo sia applicabile a qualunque tipo di livello misurabile dal fonometro. Il suo utilizzo e la sua efficacia non variano se si usano, ad esempio, livelli equivalenti o livelli massimi.

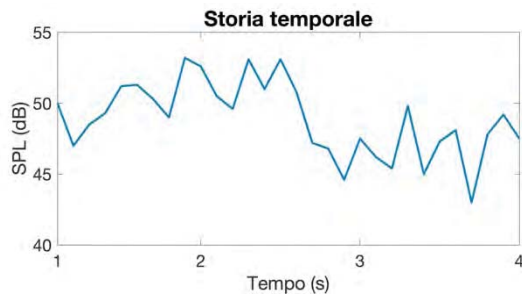


Fig. 3 – Esempio di storia temporale acquisita con un intervallo temporale di 100 ms

Example of time history obtained with an interval time equal to 100 ms

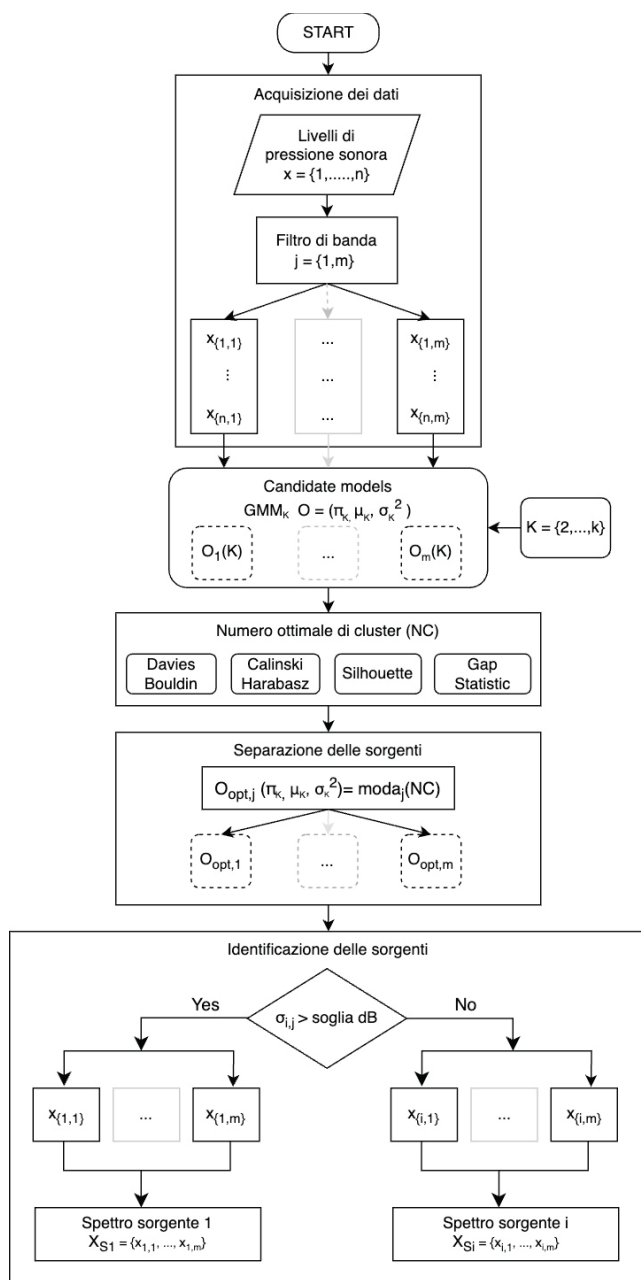


Fig. 4 – Schema a blocchi del metodo proposto. Adattato da [9]
Block scheme of the proposed method. Adapted from [9]

La ricostruzione degli spettri consiste nel raggruppare in maniera coerente tutti i cluster che hanno deviazioni standard dello stesso tipo, ad esempio umana o meccanica.

Nei casi studio descritti successivamente, sono stati utilizzati come database, cioè come popolazioni statistiche, *short-term* L_{eq} in bande d'ottava da 125 a 4000 Hz. I monitoraggi sono stati eseguiti con un intervallo di acquisizione pari a 100 ms. Questo permette di acquisire SPL anche fra le pause del parlato e dunque di evitare eccessivi mascheramenti dati dalle interazioni fra sorgenti coesistenti. In Fig. 3 è mostrato un esempio di storia temporale ottenuta con un intervallo di acquisizione di 100 ms. Qui, è possibile apprezzare le fluttuazioni del segnale. SPL più alti apparterranno ad una determinata sorgente, SPL più bassi riescono ad essere acquisiti nelle pause della sorgente più alta ed appartengono ad una sorgente differente. Questi SPL andranno dunque a disegnare le diverse forme delle occorrenze totali con i suoi diversi picchi e flessi.

In Fig. 4 è mostrato il processo del metodo proposto step-by-step attraverso uno schema a blocchi.

4 | Risultati

Di seguito si presentano i risultati suddivisi per ogni contesto di applicazione. In tutti i casi studio, le metriche di selezione del numero ottimale di cluster è sempre risultato uguale a 2. Dunque, in ogni contesto di applicazione, sono state ricercate 2 sorgenti. Nelle aule universitarie sono state separate la student activity (SA) e lo speech level (SL), negli uffici le sorgenti di rumore (di tipo meccanico e traffico) e il parlato.

4.1 | Aule universitarie

I principali risultati ottenuti nelle aule universitarie sono mostrati in Tab. 1. Qui sono indicati il codice della lezione, la corrispettiva aula, il numero di studenti presenti durante la lezione e i 3 parametri principali: la SA, l'SL e il rapporto segnale-rumore (SNR). Siccome durante tutti i monitoraggi gli impianti di condizionamento dell'aria erano spenti, l'SNR è ricavato come la differenza fra SL e SA, intendendo SA come sorgente di rumore. Le lezioni dalla A alla I sono state monitorate prima i lavori di correzione acustica, le rimanenti dopo. Il numero degli studenti sottende la percentuale di occupazione delle aule che fa variare, più o meno significativamente in base ai casi, le proprietà acustiche della stanza.

I valori medi sono stati riportati sia per il gruppo di lezioni monitorate prima che dopo i lavori. SA, SL e SNR prima dei lavori variano rispettivamente negli intervalli 47,5-61 dBA, 63,3-75,5 dBA e 13-20 dBA. Dopo i lavori, gli intervalli misurati sono rispettivamente 47,2-53,9 dBA, 59-72,1 dBA e 9,1-19,7 dBA. Mediamente è stata misurata una SA di 52,1 dBA, un SL di 68,3 dBA e un SNR di 16,2 dBA prima dei lavori. Dopo i lavori l'SA misurata è stata di 50,8 dBA, l'SL di 65,6 dBA e l'SNR di 14,9 dBA. Considerando gli outliers, l'occupazione, anche se mediamente inferiore, risulta circa la stessa.

Tab. 1 – Lezioni monitorate prima e dopo gli interventi di correzione acustica. Oltre al codice della lezione sono indicate l'aula in cui sono state svolte, il numero di studenti presenti, la SA e lo SL misurati e il corrispettivo SNR
Results of the lessons measured before and after restoration works. The lesson code, the corresponding hall of the lesson, the number of students, the SA, the SL, and the SNR are shown

	Lezione	Aula	N. studenti	SA (dBA)	SL (dBA)	SNR (dBA)
Prima	A	I	145	48,2	65,0	16,8
	B	I	200	47,5	63,3	15,8
	C	I	100	53,3	66,3	13,0
	D	I	150	51,2	67,2	16,0
	E	II	250	48,4	67,5	19,1
	F	II	160	50,3	66,5	16,2
	G	II	120	61,0	75,5	14,5
	H	II	150	55,3	75,3	20,0
	I	II	200	53,4	68,0	14,6
Media			164	52,1	68,3	16,2
Dopo	J	I	130	51,3	64,6	13,3
	K	I	185	49,9	69,6	19,7
	L	I	130	47,2	64,6	17,4
	M	I	80	49,9	59,0	9,1
	N	I	190	52,7	68,9	16,2
	O	II	110	53,9	72,1	18,2
	P	II	125	51,4	60,8	9,4
	Q	II	120	48,8	66,4	17,6
	R	II	95	52,1	64,9	12,8
Media			129	50,8	65,6	14,9

4.2 | Uffici

I principali risultati riguardanti le applicazioni del metodo proposto negli uffici si concentrano sulla possibilità di ricostruire gli spettri di ogni tipologia di sorgente presente nello spazio.

In Fig. 5 sono mostrati gli spettri ricostruiti nei due uffici oggetto di studio. Nel caso 1 (in alto) gli spettri misurati corrispondono al parlato dei lavoratori e alla somma delle sorgenti di tipo meccanico che si hanno all'interno di un ufficio (condizionamento dell'aria, computer, ecc.). Nel caso 2 (in basso), al momento della misura, l'impianto di condizionamento dell'aria era spento. Tuttavia, essendo l'ufficio a ridosso di una strada ad alto scorrimento, il rumore da traffico costituiva una sorgente rilevante durante la giornata lavorativa.

È importante notare che entrambi gli uffici presentavano caratteristiche acustiche molto differenti. Nel caso 1 il controsoffitto era completamente trattato e l'ambiente risultava, in termini di riverberazione, molto secco. Nel caso 2 l'ufficio non presentava alcun trattamento acustico.

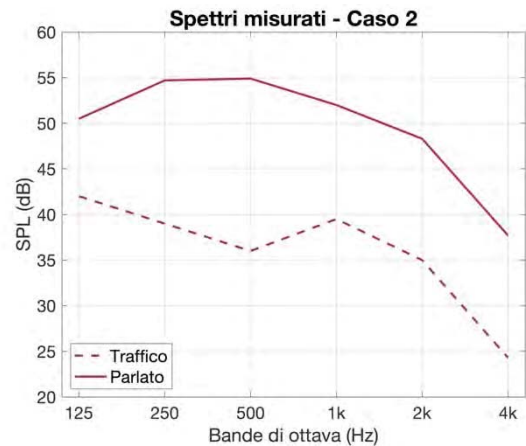
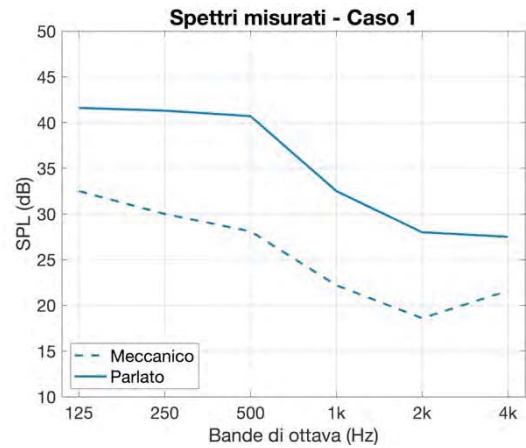


Fig. 5 – Spettri ricostruiti all'interno dei due uffici. Le linee tratteggiate indicano le sorgenti di rumore, le linee continue indicano gli spettri del parlato
Reconstructed spectra in offices. Dashed lines refer to the detected noise, solid lines refer to speech spectra

5 | Discussioni

5.1 | Aule universitarie e ruolo dell'occupazione

I risultati ottenuti all'interno delle aule universitarie mostrano come ambienti con caratteristiche acustiche conformi alla UNI11532 migliorino anche la capacità degli studenti di rimanere concentrati durante le lezioni. Al netto dell'uso dell'impianto di amplificazione, è ragionevole considerare anche uno sforzo vocale inferiore degli insegnanti.

Un approfondimento circa il cambiamento di comportamento delle persone che impiegano gli spazi misurati si può ottenere andando a correlare il numero degli studenti con i 3 livelli di pressione sonora presi in considerazione: SA, SL e SNR.

In Fig. 6 sono mostrate le correlazioni sopracitate. In tutti i grafici, il giallo indica il prima e il grigio indica il dopo lavori. I punti indicano le singole lezioni analizzate, le linee tratteggiate le corrispondenti regressioni, lineari in questo caso. Dai grafici si può intuire subito il ruolo chiave che gioca l'occupazione durante le lezioni universitarie. Nei presenti casi studio, le sedute sono disposte ad anfiteatro. In lavori precedenti sia Hodgson che Choi hanno messo in evidenza l'impatto rilevante che l'occupazione e la sua disposizione nello spazio genera sull'acustica delle aule [3,10].

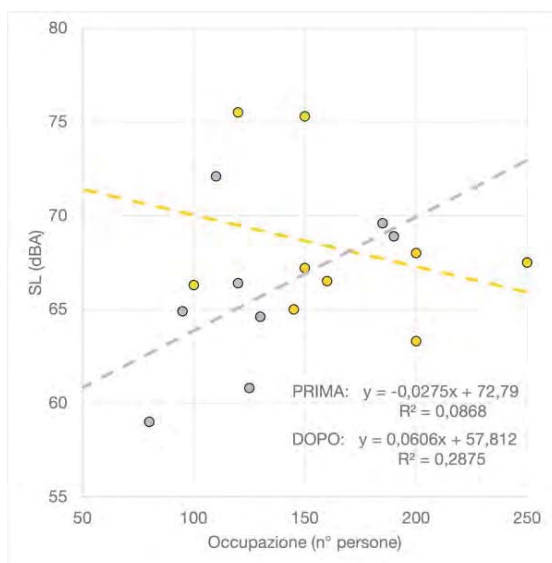
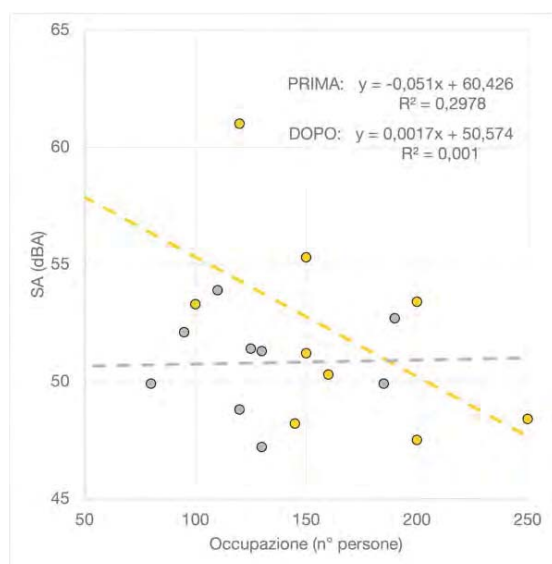
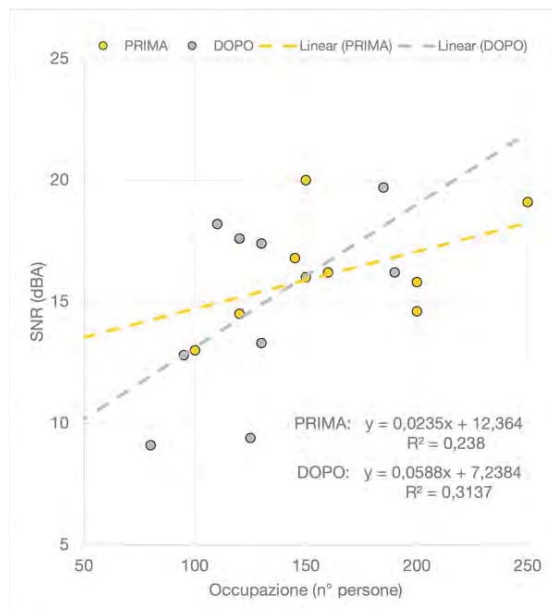


Fig. 6 – Correlazioni fra occupazione (numero di studenti durante le lezioni) e SA, SL e SNR prima e dopo i lavori di miglioramento
Correlation between occupancy (number of students attending lectures) and SA, SL, and SNR before and after restoration works

Dalla relazione fra studenti e SNR (in alto) si può notare come i lavori hanno reso la correlazione più significativa. Questo può voler dire due cose: da un lato o che più gente segue la lezione e più si fa silenzio; o che più gente segue la lezione e più l'insegnante tende ad alzare la voce o quantomeno ad avvicinare il microfono.

Nel mezzo di Fig. 6 invece è raffigurata la correlazione fra SA e numero di studenti. È possibile notare come, dopo gli interventi di correzione acustica, SA perda la correlazione con l'occupazione rimanendo costante. Questo potrebbe significare che gli studenti controllano il proprio chiacchiericcio in base al loro numero.

Nel grafico di correlazione fra occupazione ed SL (Fig. 6 in basso) sembra esserci l'effetto maggiore degli interventi di correzione acustica. Le tendenze delle regressioni cambiano completamente inclinazione suggerendo che gli insegnanti tenderebbero ad alzare il livello della propria voce all'aumentare del numero di studenti presenti in aula. Tuttavia, questo non è detto che sia strettamente legato allo sforzo vocale in quanto, come già accennato, tutte le lezioni sono svolte utilizzando un impianto di amplificazione.

Date le popolazioni statistiche di ogni scenario abbastanza piccole, i risultati numerici delle correlazioni sono da considerarsi preliminari. Essendo un confronto fra due scenari, i valori di R^2 sono riportati per evidenziare in maniera qualitativa più che quantitativa il cambio di tendenza dei dati [4].

5.2 | Uffici

Al fine di validare il metodo e di valutare l'affidabilità degli spettri misurati, è stato effettuato un confronto fra i risultati ottenuti e degli spettri di riferimento. Questo confronto è definito *spectral matching* e consiste nel valutare gli andamenti delle tipologie di spettri ottenuti con riferimenti normativi e di letteratura. In Fig. 7 sono mostrati questi confronti per entrambi i casi studio attraverso degli spettri relativi. Per spettri relativi si intendono gli spettri valutati riportando a 0 dB la banda dei 1000 Hz e valutando, rispetto a questo riferimento, le differenze delle bande rimanenti. I riferimenti di normativa o di letteratura sono indicati con linee tratteggiate mentre gli spettri misurati sono riportati con linee continue.

Negli uffici, gli spettri misurati in entrambi i casi studio sono stati suddivisi in base alla loro tipologia. Da una parte (Fig. 7 in alto) sono visualizzati gli spettri misurati di sorgenti di rumore. Nel caso specifico sono stati rilevati uno spettro di rumore meccanico e uno spettro di rumore da traffico. Per quanto riguarda il rumore meccanico, lo spettro di confronto è quello riportato dall'ASHRAE Handbook, testo di riferimento per quanto riguarda il funzionamento degli impianti di ventilazione meccanica [31]. Tuttavia, uno spettro che si riferisce al solo impianto di ventilazione risulta in questo caso relativamente limitato nel confronto. Infatti, come già accennato precedentemente, lo spettro indicato come "meccanico" in questo lavoro si riferisce all'insieme delle sorgenti di natura meccanica, e non soltanto di ventilazione. Questo spiega le

differenze rilevanti che si possono notare alle alte frequenze, cioè nelle bande dei 2000 e 4000 Hz. In quest'ultima banda, è particolarmente accentuata la differenza ed è attribuibile a ronzii di natura elettrica dei vari dispositivi presenti nell'ufficio (lampade, schermi, ecc.).

Il rumore da traffico invece è stato confrontato con lo spettro normalizzato indicato dalla EN1793-3 [32]. È importante notare che lo spettro riportato in normativa fa riferimento alla condizione di campo libero. Dunque, le proprietà acustiche dell'ufficio e della sua facciata possono influenzare i risultati. Le tendenze dei due spettri, indicati in alto della Fig.6 in rosso e giallo, è simile con la differenza più rilevante nella banda dei 500 Hz. Tuttavia, il contributo delle basse frequenze risulta coerente con il rumore prodotto da un traffico a bassa velocità [33].

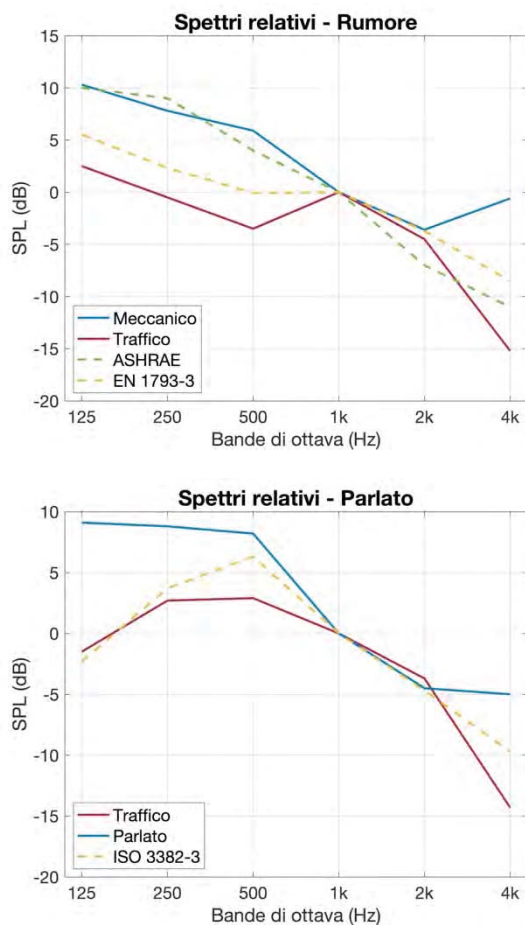


Fig. 7 – Spettri relativi e confronto con riferimenti normativi e di letteratura

Relative spectra and comparison with standards and literature

In basso della Fig. 6 invece sono mostrati gli spettri del parlato misurati in entrambi i casi studio e confrontati con lo spettro del parlato direttivo indicato dalla ISO 3382-3 [34]. Anche qui è importante sottolineare come lo spettro di riferimento sia misurato in condizioni anecoiche. Dunque, è ragionevole aspettarsi uno spettro influenzato dalle proprietà acustiche degli ambienti. In particolare, sono facilmente visibili differenze relativamente grandi alle basse frequenze. Queste

sono attribuibili al cambiamento dello spettro del parlato in ambienti riverberanti e all'incertezza di misura che si ha soprattutto ai 125 e 250 Hz [35,36]. Anche la direttività della sorgente stessa può influenzare l'andamento dello spettro rispetto al riferimento normativo. Nei casi studio abbiamo più persone che parlano, anche contemporaneamente; dunque, è ragionevole considerare una direttività complessiva diversa da quella indicata dalla ISO. In aggiunta, alle basse frequenze, anche i comportamenti modali dei due uffici possono influenzare gli spettri misurati.

6 | Conclusioni

Il presente lavoro ha proposto un metodo basato su tecniche di machine learning per separare e misurare differenti sorgenti sonore attraverso monitoraggi fonometrici. Questo metodo è stato ispirato dall'attenzione crescente, anche nel campo dell'acustica, verso il machine learning. La possibilità di sfruttare grandi quantità di dati provenienti da monitoraggi a lungo termine permette di utilizzare algoritmi sofisticati per compiere analisi più accurate e dettagliate rispetto alle tecniche usate oggi nella comune prassi tecnica. L'uso di alcune metriche, ad esempio dei livelli percentili, si basa su assunzioni temporali difficilmente dimostrabili. Inoltre, il significato fisico stesso dei livelli percentili risulta poco accurato. Il loro utilizzo è basato su alcune regole pratiche. Ad esempio, in contesti con rumori poco fluttuanti L_{50} sarà vicino a L_{eq} . Al contrario, L_{eq} sarà vicino a L_{10} e sarà maggiore di L_{90} di circa 10 dB o più. Il metodo qui proposto invece sfrutta i punti notevoli delle distribuzioni statistiche dei fenomeni misurati. Con l'utilizzo dell'algoritmo Gaussian Mixture Model è possibile separare i differenti contributi di differenti sorgenti sonore presenti nei monitoraggi a lungo termine in contesti complessi. Il metodo è stato validato e le sue assunzioni teoriche sono dimostrate in diversi lavori. L'uso dei soli livelli di pressione sonora permette inoltre di preservare la privacy degli ambienti poiché non è necessaria una registrazione audio per compiere la separazione. Dettagli del metodo e le sue applicazioni sono rimandati alla bibliografia di riferimento. Ulteriori misure permetteranno di affinare il metodo e ricavare soglie più precise per il riconoscimento delle sorgenti. L'uso del metodo proposto affiancato alla comune prassi tecnica permetterebbe ai tecnici acustici di ampliare sostanzialmente le capacità di analisi di vari scenari complessi.

Conclusions

The present work has proposed a method based on machine learning techniques to separate and measure different sound sources through sound level monitoring. This method was inspired by the growing attention towards machine learning, also in the field of acoustics. The chance of exploiting large amounts of data from long-term monitoring allows sophisticated algorithms to perform more accurate and detailed analyses compared to techniques commonly used in current technical practice. Certain metrics, such as percen-

tile levels, are based on temporally difficult-to-prove assumptions. Furthermore, the physical meaning of percentile levels is not very accurate. Their use is based on some practical rules. For example, L_{50} will be close to L_{eq} in contexts with less fluctuating noise. Conversely, L_{eq} will be close to L_{10} and greater than L_{50} by about 10 dB or more. The method proposed here instead exploits the notable points of the statistical distributions of the measured phenomena. Using the Gaussian Mixture Model algorithm, separating the different contributions of sound sources present in long-term monitoring in complex contexts is possible. The method has been validated, and its theoretical assumptions have been demonstrated in various works. Using only sound pressure levels also preserves the privacy of environments since audio recording is not necessary to perform the separation. Details of the method and its applications are deferred to the reference bibliography. Further measurements will allow for refining the method and deriving more precise thresholds for source recognition. Using the proposed method alongside common technical practices would substantially expand the capabilities of acoustic technicians to analyze various complex scenarios.

Bibliografia

- [1] IEC 61672-1:2013 | IEC Webstore, (n.d.). <https://webstore.iec.ch/publication/5708> (accessed February 20, 2024).
- [2] H. Sato, J.S. Bradley, Evaluation of acoustical conditions for speech communication in working elementary school classrooms, *The Journal of the Acoustical Society of America* 123 (2008) 2064–2077. <https://doi.org/10.1121/1.2839283>.
- [3] M. Hodgson, R. Rempel, S. Kennedy, Measurement and prediction of typical speech and background-noise levels in university classrooms during lectures, *The Journal of the Acoustical Society of America* 105 (1999) 226–233. <https://doi.org/10.1121/1.424600>.
- [4] D. De Salvio, D. D’Orazio, Effectiveness of acoustic treatments and PA redesign by means of student activity and speech levels, *Applied Acoustics* 194 (2022) 108783. <https://doi.org/10.1016/j.apacoust.2022.108783>.
- [5] K. Hummel, E. Ryherd, X. Cheng, B. Lowndes, Relating clustered noise data to hospital patient satisfaction, *The Journal of the Acoustical Society of America* 154 (2023) 1239–1247. <https://doi.org/10.1121/10.0020760>.
- [6] D. De Salvio, D. D’Orazio, M. Garai, Unsupervised analysis of background noise sources in active offices, *The Journal of the Acoustical Society of America* 149 (2021) 4049–4060. <https://doi.org/10.1121/10.0005129>.
- [7] D. De Salvio, M.J. Bianco, P. Gerstoft, D. D’Orazio, M. Garai, Blind source separation by long-term monitoring: A variational autoencoder to validate the clustering analysis, *The Journal of the Acoustical Society of America* 153 (2023) 738–750. <https://doi.org/10.1121/10.0016887>.
- [8] D. De Salvio, G. Fratoni, D. D’Orazio, M. Garai, Assessing human activity noise in workspaces using machine learning and numerical models, *INTER-NOISE and NOISE-CON Congress and Conference Proceedings* 265 (2023) 5259–5269. https://doi.org/10.3397/IN_2022_0767.
- [9] M. Cingolani, D. De Salvio, D. D’Orazio, M. Garai, Clustering analysis of noise sources in healthcare facilities, *Applied Acoustics* 214 (2023) 109660. <https://doi.org/10.1016/j.apacoust.2023.109660>.
- [10] Y.-J. Choi, Evaluation of acoustical conditions for speech communication in active university classrooms, *Applied Acoustics* 159 (2020) 107089. <https://doi.org/10.1016/j.apacoust.2019.107089>.
- [11] M.C. Cutler, M.R. Cook, M.K. Transtrum, K.L. Gee, Data-driven decomposition of crowd noise from indoor sporting events, *The Journal of the Acoustical Society of America* 155 (2024) 962–970. <https://doi.org/10.1121/10.0024724>.
- [12] D. D’Orazio, D. De Salvio, L. Anderlucci, M. Garai, Measuring the speech level and the student activity in lecture halls: Visual-vs blind-segmentation methods, *Applied Acoustics* 169 (2020) 107448. <https://doi.org/10.1016/j.apacoust.2020.107448>.
- [13] Y. Hasegawa, E. Ryherd, Clustering acoustical measurement data in pediatric hospital units, *The Journal of the Acoustical Society of America* 148 (2020) 265–277. <https://doi.org/10.1121/10.0001584>.
- [14] M.J. Bianco, P. Gerstoft, J. Traer, E. Ozanich, M.A. Roch, S. Gannot, C.-A. Deledalle, Machine learning in acoustics: Theory and applications, *The Journal of the Acoustical Society of America* 146 (2019) 3590–3628. <https://doi.org/10.1121/1.5133944>.
- [15] E. Ozanich, A. Thode, P. Gerstoft, L.A. Freeman, S. Freeman, Deep embedded clustering of coral reef bioacoustics, *The Journal of the Acoustical Society of America* 149 (2021) 2587–2601. <https://doi.org/10.1121/10.0004221>.
- [16] M.J. Bianco, S. Gannot, E. Fernandez-Grande, P. Gerstoft, Semi-Supervised Source Localization in Reverberant Environments With Deep Generative Modeling, *IEEE Access* 9 (2021) 84956–84970. <https://doi.org/10.1109/ACCESS.2021.3087697>.
- [17] W.F. Jenkins II, P. Gerstoft, M.J. Bianco, P.D. Bromirski, Unsupervised Deep Clustering of Seismic Data: Monitoring the Ross Ice Shelf, Antarctica, *Journal of Geophysical Research: Solid Earth* 126 (2021) e2021JB021716. <https://doi.org/10.1029/2021JB021716>.
- [18] J. Neri, R. Badeau, P. Depalle, Unsupervised Blind Source Separation with Variational Auto-Encoders, in: *2021 29th European Signal Processing Conference (EUSIPCO)*, IEEE, Dublin, Ireland, 2021: pp. 311–315. <https://doi.org/10.23919/EUSIPCO54536.2021.9616154>.
- [19] C.C. Aggarwal, C.K. Reddy, eds., *Data Clustering: Algorithms and Applications*, 1st edition, Chapman and Hall/CRC, Boca Raton, 2013.
- [20] C.M. Bishop, N.M. Nasrabadi, *Pattern recognition and machine learning*, Springer, 2006. <https://link.springer.com/book/9780387310732> (accessed February 20, 2024).
- [21] D.H. Wolpert, W.G. Macready, No free lunch theorems for optimization, *IEEE Transactions on Evolutionary Computation* 1 (1997) 67–82. <https://doi.org/10.1109/4235.585893>.
- [22] G.E.P. Box, D.R. Cox, An Analysis of Transformations, *Journal of the Royal Statistical Society: Series B (Methodological)* 26 (1964) 211–243. <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>.
- [23] G.J. McLachlan, S.X. Lee, S.I. Rathnayake, Finite Mixture Models, *Annual Review of Statistics and Its Application* 6 (2019) 355–378. <https://doi.org/10.1146/annurev-statistics-031017-100325>.
- [24] J. Shao, *Mathematical statistics*, Springer Science & Business Media, 2003.
- [25] X.-L. Meng, D. Van Dyk, The EM Algorithm – an Old Folk-song Sung to a Fast New Tune, *Journal of the Royal Statistical*

- Society: Series B (Methodological) 59 (1997) 511-567. <https://doi.org/10.1111/1467-9868.00082>.
- [26] T. Caliński, J. Harabasz, A dendrite method for cluster analysis, *Communications in Statistics* 3 (1974) 1-27. <https://doi.org/10.1080/03610927408827101>.
- [27] D.L. Davies, D.W. Bouldin, A Cluster Separation Measure, *IEEE Trans. Pattern Anal. Mach. Intell. PAMI-1* (1979) 224-227. <https://doi.org/10.1109/TPAMI.1979.4766909>.
- [28] P.J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, *Journal of Computational and Applied Mathematics* 20 (1987) 53-65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
- [29] R. Tibshirani, G. Walther, T. Hastie, Estimating the number of clusters in a data set via the gap statistic, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63 (2001) 411-423. <https://doi.org/10.1111/1467-9868.00293>.
- [30] UNI 11532-2:2020 – UNI Ente Italiano di Normazione, (n.d.). <https://store.uni.com/uni-11532-2-2020> (accessed February 28, 2024).
- [31] 2023 ASHRAE Handbook – HVAC Applications | ashrae.org, (n.d.). www.ashrae.org/technical-resources/ashrae-handbook/2023-ashrae-handbook-hvac-applications (accessed February 28, 2024).
- [32] UNI EN 1793-3:1999 – UNI Ente Italiano di Normazione, (n.d.). <https://store.uni.com/uni-en-1793-3-1999> (accessed February 28, 2024).
- [33] A. Can, L. Leclercq, J. Lelong, D. Botteldooren, Traffic noise spectrum analysis: Dynamic modeling vs. experimental observations, *Applied Acoustics* 71 (2010) 764-770. <https://doi.org/10.1016/j.apacoust.2010.04.002>.
- [34] UNI EN ISO 3382-3:2022 – UNI Ente Italiano di Normazione, (n.d.). <https://store.uni.com/uni-en-iso-3382-3-2022> (accessed February 28, 2024).
- [35] J.H. Rindel, C.L. Christensen, A.C. Gade, Dynamic sound source for simulating the Lombard effect in room acoustic modeling software, in: *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, Institute of Noise Control Engineering, 2012: pp. 954-966. www.odeon.dk/pdf/C118-IN12_Paper%20209_Rindel_Christensen.pdf (accessed February 28, 2024).
- [36] G. Leembruggen, J. Verhave, S. Feistel, L. Holtzem, P. Mapp, H. Sato, T. Steinbrecher, S. Van Wijngaarden, The effect on sti results of changes to the male test-signal spectrum, *Proc. IOA* 38 (2016) 78-87.