

# 近五年道路系統多音源空間聲音量測、聲源分離與辨識之研究

## 一、前言：環境噪音挑戰與研究動機

道路交通噪音已被世界衛生組織(World Health Organization, WHO)於《歐洲區域環境噪音指引》中認定為僅次於空氣污染的第二大環境健康危害因子，每年於西歐地區造成至少一百萬健康生命年(Disability-Adjusted Life Years, DALYs)的損失，其中缺血性心臟病、睡眠障礙與認知功能受損為主要的健康負擔 [1]。近五年來，多篇高品質統合分析進一步強化了此一證據基礎：Smith、Cordoza 與 Basner(2022)更新了 WHO 睡眠系統性綜述，確認夜間道路交通噪音與自述睡眠障礙間存在明確的劑量-反應關係，每增加 10 dB 之  $L_{night}$  對高度睡眠受擾者比例之增加具統計顯著性 [2]。Fu 等(2023)於 *Journal of Urban Health* 發表之劑量-反應統合分析指出，長期暴露於道路交通噪音每增加 10 dB 之  $L_{den}$ ，缺血性心臟病發生風險增加約 3–4% [3]；Chen 等(2023)於 *European Journal of Public Health* 所發表之傘狀綜述(Umbrella Review)更整合多篇系統性綜述，確認環境噪音與心血管疾病整體風險增加達 34% 有關 [4]。Perschagen 等(2025)於 *Environmental Epidemiology* 之最新研究則明確量化了道路交通噪音與心肌梗塞及中風發生率之劑量-反應關係 [5]。這些證據為全球噪音管制法規之修訂以及精確聲源分離與辨識技術之發展提供了迫切的公共衛生動機。

在此背景之下，如何精確地量測、分離並辨識道路系統中的多音源空間聲音(Multi-source Spatial Acoustic Field)，已成為當代環境噪音工程、聲學計量學與人工智慧交叉領域最為關鍵的研究命題之一。道路交通聲場的複雜性源自於多種聲源同時存在：引擎燃燒噪音(Combustion Noise)、進排氣噪音(Intake/Exhaust Noise)、動力傳動系統噪音(Powertrain Noise)、輪胎/路面交互作用噪音(Tyre/Road Interaction Noise)、空氣動力噪音(Aerodynamic Noise)，以及輔助背景環境音(鳥鳴、行人、施工等)皆在同一時空域中疊加，形成高度非定常且方向性複雜的聲場結構 [6]。

傳統上，道路交通噪音之量測主要依據國際電工委員會(International Electrotechnical Commission, IEC)所頒布之 IEC 61672-1:2013 標準，規範 Class 1 與 Class 2 噪音計(Sound Level Meter, SLM)的技術規格、頻率加權(A/C/Z-weighting)、時間加權(Fast/Slow/Impulse)與校準程序，其量測結果(如  $L_{Aeq}$ 、

$L_{den}$ 、 $L_{night}$ )為法定噪音管制與環境影響評估的核心指標 [7]。然而單點單通道的噪音計僅能提供積分式能量值，無法解析聲場的空間結構，亦無法回答「噪音來自哪一輛車、哪一類聲源」的問題，此一認知缺口催生了陣列量測(Array Measurement)、盲訊號分離(Blind Source Separation, BSS)與深度學習聲源辨識(Deep Learning-based Source Identification)等進階技術的蓬勃發展。

## 二、道路噪音之聲源組成與物理機制

在物理層次上，一輛行駛中的機動車輛之輻射聲功率可分解為兩大主要貢獻來源。其一為動力單元噪音(Power Unit Noise)，包含內燃機燃燒脈動、進排氣口輻射、曲軸傳動、齒輪箱與冷卻系統之噪音，此類聲源於低速(低於約 50 km/h)時主導整體輻射聲壓。其二為滾動噪音(Rolling Noise)，即輪胎胎塊撞擊路面所產生的胎塊衝擊音(Tread Impact Noise)、胎面花紋泵送氣柱共振(Air-pumping)以及胎體結構振動輻射，該類聲源於中高速(高於 50 km/h)時主導，且隨車速呈約  $30 \cdot \log_{10}(v)$  之對數成長關係 [6]。Braun 等學者之文獻回顧指出，於 ISO 362 之加速通過試驗(Accelerated Pass-by Test)中，輪胎/路面噪音貢獻量可於現代乘用車達整體輻射音能之 60%至 75%，成為降噪工程最關鍵且最難單獨量化之目標 [6]。

為分離量測輪胎/路面噪音，聲學界發展出兩種國際廣泛採用的方法：其一為近接法(Close-ProXimity method, CPX method)，依據 ISO 11819-2 規範於拖車內艙以固定距離(通常為 20 cm)的麥克風陣列追蹤兩支參考胎之輻射音；其二為統計通過法(Statistical Pass-By method, SPB)，依 ISO 11819-1 規範於路側 7.5 m、高度 1.2 m 處量測大量實車通過之最大 A 加權聲壓級  $L_{AFmax}$ ，並依速度回歸得到對應車速下之參考噪音水平 [8]。de León 等(2020)之研究進一步將 CPX 方法與路面紋理量測結合，證實路面紋理光譜與 CPX 噪音光譜間具高度可重現之因果關聯，該方法現已成為歐盟低噪音路面(Low-Noise Pavement)鑑定之法定技術 [8]。Li(2022)於香港理工大學之博士論文系統性地推進了 CPX 方法學之精度與不確定度量化框架 [9]，而 Baumgartner、Fuchs 與 Haider(2023)更於 Forum Acusticum 會議證實，透過機器學習演算法關聯不同 CPX 試驗間之量測差異，可顯著降低實驗室間之再現性偏差 [10]。Leung、Rai 與 Li(2024)則拓展了 CPX 方法於滾動胎聲功率輻射之應用範疇 [11]。

## 三、歐盟 CNOSSOS-EU 與國際法定量測框架

於歐盟層級，指令(EU) 2015/996 附錄所載之 CNOSSOS-EU(Common Noise Assessment Methods in Europe)方法現已成為歐盟各成員國履行「環境噪音指令」

(Environmental Noise Directive, 2002/49/EC)製作策略性噪音地圖(Strategic Noise Maps)之唯一合法計算方法。該方法將道路聲源分為「動力單元」與「輪胎/路面」兩個半對應實質物理機制的排放組件，並以車輛分類(輕型車、中重型車、重型車、機動二輪車、電動車)與交通流量參數計算線源發射位準 [12]。Peeters 與 van Blokland(2018)指出原始 CNOSSOS-EU 公式於低速段(<30 km/h)與特定車類存在顯著高估或低估，已由歐洲學界提出修正係數並於後續修訂中部分納入 [12]。Martinez 與 King(2026)於 Acta Acustica 期刊所發表的最新歐洲道路噪音資料透明化研究中，進一步呼籲建立統一且可審計的資料表達規範，以支援跨國健康負擔分析與公眾溝通 [13]。

在 AI 預測模型的近期發展方面，Umar 等(2024)於 Cogent Engineering 發表之系統性書目綜述涵蓋了人工智慧模型於交通噪音預測之整體發展脈絡，分析指出交通流量、車速、車種組成與距離為幾乎所有機器學習模型共同採用之核心輸入變數，而於低速都市情境中，機器學習模型對 CNOSSOS-EU 之精度優勢可達 2-4 dB [14]。Goncharenko(2025)於芬蘭 Tampere 與 Oulu 兩城市之實證研究進一步證實，利用時序交通流量資料估計噪音水平可大幅降低實地量測成本，惟模型於主幹道與支線道路間之可遷移性仍有限 [15]。Kumar 等(2025)於印度 Ranchi 之田野研究則將機器學習分類器應用於噪音煩擾度預測，顯示交通噪音對住民心理健康之影響於中低收入國家同樣顯著，機器學習可作為地方性環境健康治理之務實工具 [16]。

在法定量測端，ISO 362-1/-2 規範乘用車加速通過聲級測試、ISO 10844 規範試車道表面特性、ISO 11819-1/-2 規範路面交通噪音、而 ISO 1996-1/-2 則為社區噪音環境量測的主流標準。Braun 等(2013)指出，ISO 362 雖為全球車輛型式認證之核心程序，但其量測結果與實際都市交通噪音水平之相關性仍有限，主因在於該試驗代表單車加速之最壞情境，而非多車流穩態行駛之都市聲學現實 [6]。於非歐盟國家，美國聯邦環保署(USEPA)、日本環境省與中華民國環境部皆採用類似 IEC 61672 Class 1/Class 2 噪音計作為法定計量工具，並以  $L_{eq}$ (等效連續音壓級)、 $L_{max}$ 、 $L_{10}$ 、 $L_{90}$  等統計量為管制指標。

#### 四、多音源空間聲音量測方法：技術譜系與近期進展

在空間聲學量測層次上，可分為幾個技術層級。最基礎之單點全向性麥克風(Omnidirectional Microphone)僅能捕捉標量聲壓，無法提供方向資訊；進階之聲強探棒(Sound Intensity Probe, p-p 或 p-u 型)可提供向量聲強資訊，適用於建築

隔音與機械聲源功率量化；聲照像 (Acoustic Camera) 與波束成形陣列 (Beamforming Array) 則可將聲源於二維空間上成像，特別適用於移動車輛通過試驗。Michel(2006)之回顧指出，早在 1990 年代初期日本學者 Fujita 即開發出二維麥克風陣列用於高速移動車輛之聲源定位，此為現代聲照像技術之濫觴 [17]。近五年來，波束成形於車輛通過噪音量測之應用取得實質進展。Czuchaj 與 Sarradj(2024) 於 Berlin Beamforming Conference 提出以移動焦點波束成形 (Moving Focus Beamforming) 自麥克風陣列資料獲得可靠車輛通過聲源功率位準之新方法，該方法可將 ISO 362-1:2022 所要求之最大通過位準估計不確定度控制於  $\pm 1.5$  dB 內，對於車輛型式認證之輔助診斷具有重要意義 [18]。Knappe 等 (2023) 則發展 Helmholtz 反向波束成形 (Helmholtz Inverse Beamforming) 用於局部聲源估計並應用於通過噪音虛擬化 (Pass-by Noise Virtualization)，為汽車產業提供車輛開發階段之虛擬試驗平台 [19]。Fredianelli 等 (2023) 進一步比較了不同波束成形演算法於港口環境船舶與交通源之定位表現，指出反卷積類方法 (DAMAS、CLEAN-SC) 於低頻段 ( $< 500$  Hz) 之旁瓣抑制明顯優於傳統延遲相加法 [20]。Pinel 等 (2023) 則於 Forum Acusticum 提出整合聲學成像與機器學習於實地車輛通過噪音量測之流程，可將通過位準校正至 ISO 362-1:2022 標準條件下，為聲學成像與法定量測之橋接提供範例 [21]。Kron 等 (2025) 則展示如何從波束成形資料建立可聽化 (Auralization) 用之移動車輛聲源模型，為環境影響評估之感知驗證 (Perceptual Validation) 開啟新路徑 [22]。

高階 Ambisonics (Higher-Order Ambisonics, HOA) 為另一重要方向，其透過球諧函數 (Spherical Harmonic Functions) 展開聲場，能以有限麥克風數 (如 A-format 四通道或高階 32 通道) 重建全向性聲場並進行後續空間濾波與雙耳化渲染。Rafaely 等 (2022) 於 Acta Acustica 發表之綜述論文指出，Ambisonics 結合參數化空間音訊處理已成為都市聲景 (Soundscape) 研究與虛擬實境還原之重要工具，然而其空間解析度仍受球諧截斷階數限制 [23]。Georgiou 等 (2024) 進一步驗證了採用 HOA 與 360° 全景視訊結合之戶外聲景記錄方法學，可精確還原野外聲景的視聽一致性，為環境聽覺實驗與感知研究提供嚴謹的刺激建構範式 [24]。

近年微機電系統麥克風陣列 (MEMS Microphone Array) 的成熟大幅降低了陣列量測的成本。Pecioski 等 (2024) 於 INTER-NOISE 會議展示了一套基於 MEMS 的低成本都市環境聲源定位系統，證明即便採用消費級組件，於信噪比高於 10 dB 之都市交通場景中仍可達合格的聲源方位辨識精度 (約  $\pm 5^\circ$ ) [25]。Chakravarthula

等(2023)更進一步於 CVPR 頂級會議發表了長距離聲學波束成形結合神經網路之多模態場景理解系統，可於數十公尺外同時定位並成像道路上的主動聲源，展示了感知人工智慧(Perceptual AI)與空間聲學之深度融合 [26]。Zhang 等(2026)於 ACM Transactions on Sensor Networks 發表了專屬車輛聲學定位資料集與結合濾波策略，為後續研究提供了可重現的基準 [27]。

#### 五、聲源分離方法：從盲訊號分離到深度學習端到端模型

聲源分離(Source Separation)為從混合錄音中恢復個別音源訊號的逆問題。傳統以獨立成分分析(Independent Component Analysis, ICA)為代表之盲訊號分離假設源訊號彼此統計獨立，Saruwatari 等(2003)將子頻帶 ICA 與空域零波束成形(Null Beamforming)結合，顯著改善了卷積混合情境下的分離收斂性，此類混合法至今仍為麥克風陣列分離之基準 [28]。非負矩陣分解(Non-negative Matrix Factorization, NMF)則成為單通道分離之主流：Ozerov 與 Févotte(2009)提出多通道 NMF 於卷積混合環境之推廣，其在房間迴響條件下仍可有效分離立體聲音源，該論文被引用超過 850 次，為該領域之經典文獻 [29]。Févotte、Vincent 與 Ozerov(2018)後續進一步系統化了 NMF 於單通道聲源分離中的發散函數( $\beta$ -Divergence)、約束條件與演算法族譜 [30]。NMF 類方法對道路交通情境之應用包含將胎噪(寬頻隨機性)、引擎階次噪音(諧波結構)、背景風切聲分解至不同頻譜字典。

深度學習聲源分離自 Luo 與 Mesgarani(2019)所提出之 Conv-TasNet 起，於 2020 年後以時頻遮罩估計(Time-Frequency Masking)與波形域端到端模型取得突破性進展，該模型被引用達兩千九百餘次，為該領域之里程碑 [31]。近五年來，多項改進型模型陸續發展：Ravenscroft、Goetze 與 Hain(2022)提出之 Att-TasNet 於時域音訊分離中加入注意力機制，顯著改善了含噪迴響語音混合之分離表現 [32]；Shi 等(2024)之改進型 Conv-TasNet 進一步縮短了訓練時間並提升於混響環境下之尺度不變信噪比(Scale-Invariant SNR, SI-SNR)達 2 dB [33]；Ansari 等(2023)於 IEEE Access 發表之混合神經網路架構結合 U-Net 與注意力機制，專門處理交通混合聲中的語音分離問題，於實車輛內噪音情境中將 SI-SNR 改善達 12 dB 以上 [34]。Xie 與 Zhao(2025)將盲訊號分離與殘差神經網路結合，處理車內多人語音情感辨識問題，為車艙智慧化提供方法路徑 [35]。Yin 等(2025)於 ICASSP 2025 發表之文本查詢式聲源分離結合事件偵測模型，為都市聲景中按語意搜尋特定聲源(例如「只聽救護車」或「只聽摩托車」)提供了突破性工具，

此類方法極具法醫聲學(Forensic Acoustics)與智慧監測應用潛力 [36]。Ni 與 Zhou(2025)於 Scientific Reports 則將盲訊號分離結合注意力機制應用於無人機聲學分類，顯示此類方法於跨領域移動聲源之可遷移性 [37]。Jagatheesaperumal 等(2023)將盲訊號分離結合深度學習應用於智慧城市的路面品質評估，證實聲學特徵可反映路面粗糙度與損壞型態，為路面管理決策提供新工具 [38]。

#### 六、聲源辨識與分類：從手工特徵到預訓練大模型

聲源辨識(Source Identification / Classification)關注於判定混合訊號中存在哪些聲源類別。UrbanSound8K、ESC-50、AudioSet 等基準資料集的建立使得深度學習方法可在標準化條件下比較。Piczak(2015)之卷積神經網路(CNN)模型為環境聲音分類開創了深度學習時代，其被引用達一千四百餘次 [39]。Demir 等(2020)提出之新型深度 CNN 在 UrbanSound8K 上達到領先的分類正確率 [40]，Mushtaq 與 Su(2020, 2021)則提出以 Mel 頻譜圖像結合資料增強之正則化深度 CNN，於 ESC-10、ESC-50 與 UrbanSound8K 取得進一步提升 [41, 42]。

近五年最重要之範式轉變為預訓練大型音訊模型(Pretrained Audio Foundation Models)的興起。Kong 等(2020)所提出之 PANNs(Pretrained Audio Neural Networks)於 AudioSet 二百餘萬筆標註資料上預訓練，至 2026 年累計引用已達一千九百餘次，成為環境音辨識之基礎模型 [43]。然而 PANNs 原始模型參數量龐大，不利於邊緣部署，因而促成 Singh、Liu 與 Plumbley(2023)所提出之 E-PANNs，透過模型壓縮與知識蒸餾將 PANNs 縮小至原參數量的四分之一，同時保留 AudioSet mAP 達 0.42，為邊緣部署奠定基礎 [44]。Choudhary 等(2022)之 LEAN(Light and Efficient Audio classification Network)進一步將參數量縮減至 PANNs 的十分之一，同時於 YAMNet 基線上保持競爭力 [45]。

於道路交通與都市聲景之具體應用方面，多音源事件偵測(Polyphonic Sound Event Detection, SED)為更具挑戰性之任務，需同時偵測多個重疊事件之類別、起始與結束時間。Chan 與 Chin(2020)之綜述完整梳理了此領域從隱馬可夫模型至 Transformer 自注意力機制的發展脈絡 [46]。Pankajakshan 等(2019)以多任務學習同時處理 SED 與聲音活動偵測(Sound Activity Detection)，解決了能量掩蔽(Masking)對低能量事件的偵測問題 [47]。Shabbir 等(2024)於 IEEE Access 以堆疊式集成深度學習偵測都市聲景中的緊急車輛鳴笛，於巴基斯坦 Islamabad 實地收集的資料集上達到 99% 分類正確率，顯示深度學習已成熟至可實際部署於智慧交通管理基礎設施 [48]。Reza 等(2025)於 IEEE Transactions on Intelligent

Transportation Systems 發表之多頭注意力 Transformer 結合時序卷積網路(TCN) 模型專門處理道路交通事件監測，相較於傳統 CRNN 模型提升了 F1 分數達 8%，並於 UrbanSound8K 及列車到站偵測任務上皆展現強勁效能 [49]。Hvastja 等(2025)於 Heliyon 期刊發表之研究將深度學習應用於車輛燃料類型(汽油、柴油、油電混合、純電動)之聲學分類，為同時評估空氣與噪音污染提供一體化方法路徑，於塞爾維亞實地資料集上達 89% 燃料類型分類正確率，此研究對於電動車普及下之都市聲景演變監測尤具價值 [50]。Cai 等(2024)之 MAT-SED(Masked Audio Transformer for SED)模型以自監督預訓練提升聲事件偵測之標記效率，為資料稀缺之都市聲景應用提供新範式 [51]。針對都市噪音專屬之應用，Alsouda、Pllana 與 Kurti(2019)於物聯網(IoT)框架下比較了支援向量機(SVM)、K 近鄰(KNN)、Bagging 與隨機森林(Random Forest)於都市噪音辨識之效能，確立低功耗邊緣裝置上的可行方法路徑 [52]。Teng 等(2024)則將都市噪音資料與地理空間資料結合，以機器學習方法建立都市噪音舒適性評估模型，展現了從單純分貝量測邁向語意式都市聲景治理的範式轉變 [53]。

#### 七、聲源事件定位與偵測(SELD)與自監督空間聲學學習

聲源事件定位與偵測(Sound Event Localization and Detection, SELD)進一步將方向到達角(Direction of Arrival, DoA)估計與事件分類整合為單一端到端任務。DCASE(Detection and Classification of Acoustic Scenes and Events)Challenge 自 2019 年引入 SELD 任務後，於 2023–2024 年進一步加入距離估計(Distance Estimation)與多模態融合。Yeow 等(2024)所提出之 Squeeze-and-Excite ResNet-Conformer 架構於 DCASE 2024 Task 3 加入了聲源距離估計維度，為三維空間聲景分析提供完整方案 [54]。Jiang 等(2024)則於 ICASSP 會議發表了視聽融合(Audio-Visual Fusion)SELD 方法，於 DCASE 2023 挑戰中排名前列，驗證了多模態資訊融合於低資源真實情境之有效性 [55]。Dong 等(2025)於 ICASSP 2025 進一步提出聯合建模 SELD 與聲源距離估計之方法，展示了三維 SELD 任務之最新進展 [56]。Zhang 等(2024)提出之多重注意力融合 ResNet 於多聲源環境中之 SELD 效能亦獲肯定，特別適用於都市多聲源環境 [57]。

自監督學習(Self-Supervised Learning, SSL)成為解決空間聲學標註資料稀缺問題的重要方向。Yang 與 Li(2024)於 IEEE/ACM TASLP 發表之自監督空間聲學表徵學習方法，以跨通道信號重建(Cross-channel Signal Reconstruction)與多通道 Conformer 學習空間聲場特徵，為標註資料稀缺條件下之空間聲學學習提供新框

架 [58]。Santos 等(2025)更進一步探索從單聲道錄音生成一階 Ambisonics(First-Order Ambisonics, FOA)之自監督學習方法，為大規模空間聲學資料缺乏之問題提供可擴展解方 [59]。都市聲景多標籤標記方面，Cartwright 等(2019)建立之 SONYC-UST(Sounds Of New York City Urban Sound Tagging)資料集持續作為多標籤都市聲分類之基準 [60]；Bai、Chen 與 Wang(2022)於 IEEE Transactions on Cognitive and Developmental Systems 提出之時空上下文多模態都市聲標記方法進一步整合聲學與地理位置資訊，為都市聲景智慧治理提供範式 [61]；Tailleur 等(2023)於 DCASE Workshop 發表之 Spectral Transcoder 探索於欠取樣頻譜表徵上應用預訓練都市聲分類器之新路徑 [62]；Tailleur(2025)之博士論文則系統性整合了大規模資料驅動聲源監測、都市聲景分析、視覺化與摘要之完整框架 [63]。

#### 八、邊緣計算、TinyML 與大規模感測器網路部署

大規模都市聲學感測器網路之部署需求催生了輕量化模型與低功耗硬體之研究浪潮。Bellafkir、Vogelbacher 與 Freisleben(2024)於 ICSOC 會議報告了於資源受限邊緣裝置上之都市聲分類實作，將 EfficientNet-B0 與 MobileNetV3 量化至 8 位元整數後仍保留 UrbanSound8K 基準上 80% 以上之正確率 [64]。Grau-Haro 等(2025)於 arXiv 發表之綜合評估研究則測試了所有 PANNs 架構於 Raspberry Pi 平台之推論效能，為實地感測器部署提供具體硬體-模型選型指引 [65]。Munirathinam 與 Vitek(2026)於 IEEE Access 發表之 CNN-GRU 混合比例優化研究針對資源受限音訊分類提出從參數效率到微控制器(Microcontroller Unit, MCU)部署之系統性方法論 [66]。Zakaria、Amr 與 Ragheb(2025)於 IEEE Access 發表之智慧都市 IoT 綜述則將聲學感測器納入完整智慧城市基礎設施框架中檢視 [67]。

#### 九、道路系統多音源量測、聲源分離與辨識之完整技術比較表

就道路系統而言，「多音源空間聲音量測」、「聲源分離 (Source Separation)」與「聲源辨識 (classification / identification)」其實是三個層次不同、但必須串接思考的技術問題。第一層是量到多少聲音，也就是法定量測最重視的聲壓級、均能音量、統計音量與頻譜結果；第二層是把混在一起的聲音拆開，釐清不同車種、不同車道、不同高度或不同設施的貢獻；第三層才是進一步說明那個被拆出的聲音「是什麼」，例如是機車、重車、輪胎與路面接觸噪音、引擎噪音、橋面振動再輻射噪音，或道路與軌道複合場景中的不同運具來源。現行國際與我國法制的核心，仍然建立在符合標準之聲級計與規範化量測程序之上；但陣列麥克風、聲

學相機、聲源分離與深度學習辨識，已經非常適合做「診斷、歸因、治理優先排序與科技執法輔助」，只是原則上不能逕自取代法定聲級值本身。

綜合上述法定量測、陣列技術、分離演算法與辨識模型之發展，以下表 1 列本論述之核心技術譜系，並特別擴充「是否可用於法定量測」欄位與「聲源分離(Source Separation)」、「聲源辨識(Classification / Identification)」之完整技術對照：

以下將比較表 2 擴充為較完整版本。表中的「是否可用於法定量測」分成四種判讀：一是「可直接作為法定主體量測」；二是「可作為法定量測之輔助模組」；三是「可作為行政佐證或科技執法輔助」；四是「目前原則上不宜直接作為法定主體量測」。這樣分類的原因在於，法定量測要的不只是技術上可行，而是儀器性能、檢定／校正、測點幾何、動特性、測時規則、背景音處理、重現性與不確定度表達都必須能納入既有規範體系。

從法定量測的角度看，最重要的原則不是「哪一種技術最先進」，而是「哪一種技術的輸出量，可以穩定對接現行法規中的受管制量」。ISO 1996-2 的目的，是決定作為環境噪音評估基礎的聲壓級；IEC 61672-1、61672-2、61672-3 則分別界定聲級計的性能、型式評估與定期檢測程序；我國《環境音量標準》亦明確要求道路環境音量測定須使用 CNS 7129 一型噪音計或 IEC 61672-1 Class 1 噪音計，並規定測高、測點、動特性與測定時間。因此，若輸出結果是法定  $L_{eq}$ 、 $L_{max}$ 、 $L_{10}$  或 1/3 八音階等標準聲學量，且儀器與程序可檢核，它就能進入法定主體量測；反之，若輸出是「熱圖」、「方位角」、「來源機率」、「分類標籤」或「分離後的相對成分」，那通常就屬於輔助診斷或佐證體系，而不是直接替代法定值。

表 1

技術方法	核心原理	空間解析能力	聲源分離能力	聲源辨識能力	典型精度/動態範圍	成本	是否可用於法定量測	代表文獻
Class 1 積分型噪音計	單通道 A 加權積分 L <sub>eq</sub>	無(單點標量)	無	無	±1.1 dB (IEC 61672)	低~中	<input checked="" type="checkbox"/> 是 (IEC 61672-1:2013 之 Class 1 為國際法定計量器具)	[7]
ISO 362 加速通過試驗	車輛加速下路側 7.5 m 量測 L <sub>AFmax</sub>	路側單點	無(僅總量)	無	法定認證級	中	<input checked="" type="checkbox"/> 是(全球車輛型式認證法定程序)	[6]
ISO 11819-1 SPB 統計通過法	大樣本實車通過統計回歸	路側單點	間接分離(依速度回歸區分胎噪與動力噪)	車類分類(輕/重車)	±1 dB(樣本足夠)	中	<input checked="" type="checkbox"/> 是(歐盟路面認證法定方法)	[8]
ISO 11819-2 CPX 近接法	拖車內距輪胎 20 cm 麥克風陣列	輪胎局部	可分離輪胎/路面音	胎/路面型態	±0.5~1 dB	中~高	<input checked="" type="checkbox"/> 是(歐盟低噪音路面認證)	[8, 9, 10, 11]
CNOSSOS-EU 計算模型	半實證排放模型(動力+滾動兩組件)	線源/網格	模型層次分離	5 類車輛	±2~3 dB(校正後)	中	<input checked="" type="checkbox"/> 是(歐盟 EU 2015/996 法定計算方法)	[12, 13, 14]
聲強探棒(p-p / p-u)	量測聲強向量	局部向量	小範圍方向性分離	無	±1~2 dB	中	<input type="checkbox"/> 部分 (ISO 9614 聲功率測定法定)	-
波束成形陣列 (Beamforming)	延遲相加/反卷積成像	二維(高)	空間濾波分離	結合 AI 可辨識	±2~3 dB	高	<input checked="" type="checkbox"/> 否(目前為診斷工具, 輔助 ISO 362 認證)	[17, 18, 19, 20, 21, 22]

技術方法	核心原理	空間解析能力	聲源分離能力	聲源辨識能力	典型精度/動態範圍	成本	是否可用於法定量測	代表文獻
Ambisonics 球諧陣列	球諧函數展開重建聲場	全向向量	參數化空間分離	結合深度學習類	方向 $\pm 5\sim 10^\circ$	高	✗ 否(為聲景研究與沉浸式應用工具)	[23, 24]
MEMS 低成本陣列	多數位 MEMS + DSP	二維(中)	波束成形/TDOA 分離	結合分類器	$\pm 3\sim 5$ dB	低	✗ 否(不符 IEC 61672 Class 1/2)	[25, 27]
ICA / 盲訊號分離(BSS)	統計獨立性最大化	依陣列而定	強(卷積/瞬時混合)	無(僅分離)	SDR 5~15 dB	低~中	✗ 否(研究/後處理工具)	[28, 34, 37]
NMF 非負矩陣分解	頻譜字典線性組合	無或弱	強(單/多通道)	中(字典比對)	SDR 3~12 dB	低	✗ 否(研究/工程分析工具)	[29, 30]
Conv-TasNet / Att-TasNet	時域端到端分離	依陣列	極強(SI-SNR 12 dB+)	需結合下游分類器	SI-SNR 10~18 dB	中	✗ 否	[31, 32, 33]
文本查詢式分離+SED	語意條件分離	時序	語意驅動分離	強(條件式偵測)	依查詢精度	高	✗ 否	[36]
CNN 聲音分類 (UrbanSound8K)	梅爾頻譜圖+CNN	無	無	強(70~98%)	F1 0.7~0.95	中	✗ 否(研究/監測輔助工具)	[39, 40, 41, 42]
PANNs / E-PANNs / LEAN	AudioSet 預訓練大模型	無	無	極強(mAP 0.42+)	F1 0.8~0.95	中	✗ 否	[43, 44, 45]
傳統 ML(SVM/KNN/RF)	手工特徵+淺層分類器	無	無	中(60~85%)	F1 0.6~0.85	低	✗ 否	[52]
SED 多聲源事件偵測	CRNN/Transformer/MAT-SED	時序	時間域弱分離	強(多標籤)	F1 0.5~0.8	中~高	✗ 否(研究工具)	[46, 47, 51]

技術方法	核心原理	空間解析能力	聲源分離能力	聲源辨識能力	典型精度/動態範圍	成本	是否可用於法定量測	代表文獻
Transformer+TCN 道路事件	多頭注意力+時序卷積	時序	弱	強 (99% 鳴笛)	F1 0.9+	中~高	✗ 否(輔助監測)	[48, 49]
車輛燃料類型聲學分類	CNN + 聲學特徵	無	無	強 (89% 準確)	F1 0.89	中	✗ 否(空噪共管輔助)	[50]
SELD 聲源事件定位+偵測	CRNN + DoA 回歸	二維/三維	時空分離	強(類別+位置)	F1 0.4~0.7, DoA $\pm 10^\circ$	高	✗ 否	[54, 55, 56, 57]
自監督空間聲學表徵	Conformer + 跨通道重建	全向	特徵層次分離	可遷移至下游任務	依下游任務	高	✗ 否	[58, 59]
長距離神經波束成形	波束成形+深度學習	遠距高解析	強	強(多模態融合)	景深 $\pm 20$ m	很高	✗ 否	[26]
邊緣 AI / TinyML 聲學	量化 CNN + MCU	無或點	無	中~強 (80%+)	F1 0.75~0.90	低	✗ 否(可配合法定感測)	[64, 65, 66]
複合管線(陣列+BSS+SED+分類)	多階段融合	全向高解析	極強	極強	綜合最佳	極高	⚠ 未來可作為法定輔助工具	整合 [17-67]

表 2

技術或方法	主要功能定位	空間資訊能力	聲源分離能力	聲源辨識能力	面對道路多車併行/複合音源之適用性	典型輸出	是否可用於法定量測	在道路系統中的最佳角色	主要限制
Class 1 噪音計	法定聲級量測	低	無	低	可量總量，但難分離來源	dB(A)、Leq、Lmax、頻譜	可直接作為法定主體量測	法定判定、基準值、申訴處理	只能量「總暴露」或特定測點值，無法自行完成來源拆解
固定式環境噪音監測站	長期趨勢監測	低	無	低	可掌握時段變化，不擅長複合來源歸因	長期時序、統計量、頻譜	可作法定或準法定監測主體，視制度而定	趨勢管理、熱點偵測、政策評估	對單一來源貢獻不敏感
單一麥克風事件分析	事件偵測初步分類	很低	低	中	對稀疏事件尚可，對高度重疊場景較弱	波形、頻譜、事件時間戳	原則上不宜直接作為法定主體量測	低成本前端監測	容易受背景音、重疊音、反射音影響
雙通道或小型陣列之到達角估測 (TDOA/DoA)	粗略方向判別	中	低至中	低	在單主源或稀疏車流可用，多車重疊時快速退化	方位角、到達時間差	可作法定輔助，不宜單獨作法定主體	判斷噪音來向、排除非目標方向干擾	空間解析度受基線、頻率、幾何配置限制
麥克風陣列波束成形 (Beamforming)	空間定位與熱點成像	高	中至高	低至中	對移動車輛、車道分離、部件熱點分析很有效	聲源熱圖、方位、相對能量	原則上屬法定輔助，不取代聲級計	複合音源定位、工程改善診斷	多為相對量，不一定直接等於法定聲級；低頻解析與旁瓣問題明顯
聲學相機 (Acoustic Camera)	波束成形視覺化	高	中至高	中	對多車道與移動目標很實用	影像疊合聲圖、頻帶熱圖	可作行政佐證；特定制度下可結合執法	科技執法、源頭熱點可視化	需要相機同步、視線條件、天候與演算法校核
盲源分離/矩陣分解 (ICA/NMF 等)	混合訊號拆解	中 (依感測器配置)	中	中	對部分重疊聲源可改善分類前處理	分離後聲軌、成分譜	原則上不宜直接作法定主體量測	作為前處理，提高後端分類或歸因能力	對訓練資料、源數估計與場景穩定性敏感
深度學習聲源分離	非線性混合拆解	中至高	高	中	對複雜重疊場景最有潛力	分離波形、遮罩、來源機率	原則上不宜直接作法定主體量測	未來複合場景拆解源核心工具	可解釋性、資料集偏差、跨場景泛化仍是瓶頸
聲音事件偵測與分類 (SED/ASC/CNN-LSTM 等)	「是什麼聲音」的辨識	低至中	低到中	高	在有良好前處理時可辨識車種、喇叭、引擎、煞車等	類別標籤、時間戳、置信度	可作輔助佐證，不宜直接取代法定量測	自動標註、源別統計、智慧治理	辨識正確不代表可直接作為法規責任歸屬
音訊+影像融合 (聲音照相、音視同步)	定位、辨識與對象綁定	高	中至高	高	對多車同時通過與執法對象關聯最實用	車牌、影像、時序聲圖、超標事件	在特定制度設計下可作科技執法依據	車輛噪音執法、證據鏈補強	需法規授權、資料保全、同步精度與排除誤判機制
聲學向量感測器 (AVS)	單點方向資訊量測	中	中	中	在低頻方向判別與被動監測有潛力	聲壓+粒子速度估計	目前多屬研究或輔助用途	未來低頻或智慧路側感測	制度標準化、場域校正與法律位階尚不足
分散式聲學感測 (DAS)	大範圍交通感知	中到高	中	中	對長距離線狀道路監測具潛力	時空分布、事件軌跡	目前原則上不宜直接作法定主體量測	區域級監測、交通與噪音聯合分析	還在快速發展，法定標準與聲學對應關係未成熟

也就是說，真正穩健的制度設計不應把「法定量測」與「複合音源診斷」混為一談，而應採雙軌架構：第一軌仍用 Class 1 噪音計取得可執法、可申訴、可重現的標準聲級；第二軌再用麥克風陣列、聲學相機、盲源分離、深度學習辨識等方法去回答行政管理最在意、但傳統噪音計難以回答的問題，例如哪一個車道、哪一種車輛、哪一個結構部件、哪一個時段，對受體端噪音暴露貢獻最大。這種分工其實最符合現代噪音治理需求，因為法規需要穩定，治理需要精細；若硬把還在快速演進的分離或辨識模型直接塞進法定主體量測，不但容易引發儀器等效性、演算法透明性、不確定度與訴訟攻防問題，也會讓制度失去操作穩定度。

若從「道路系統多音源空間聲音量測」來看，傳統單點噪音計最大的優點是結果可直接對接法規，但最大的限制則是它只能在測點上量到「總和」，不能天然拆出不同來源。對於單一路段、單一主音源、交通組成相對單純的場景，這不是大問題；但對多車道、高架與平面併存、交叉口加減速、橋面與路面混合、甚至道路與鐵路相鄰的複合場景而言，總量雖然可以判斷是否超標，卻很難判斷治理責任與改善優先順序。這也是為何 CNOSSOS-EU 與歐盟環境噪音管理制度一方面強調指標與評估一致性，另一方面又需要更細緻的模型與補充觀測去處理不同道路源、交通條件與空間分布的差異。

麥克風陣列與波束成形的價值，正是在這裡開始顯現。Chiarotti 等人的綜述指出，波束成形本質上是利用感測器陣列進行空間濾波，藉由數學上「轉向」陣列關注某一方向或某一位置，進而完成聲源定位；而聲學相機只是把這種定位結果與光學影像耦合，使其可視化。針對移動聲源，Michel 等人的工作更直接指出，相位式麥克風陣列已成為研究道路車輛、鐵路車輛與飛機移動聲源的重要方法，因為它能在目標運動中辨識主要發聲部位，如輪胎—路面噪音、輪軌噪音或氣動噪音。對道路系統來說，這類方法最強的不是產生一個更「法定」的 dB 值，而是讓管理者看見「哪裡在吵、哪一帶最強、不同頻帶是哪個部位在主導」。因此，若目的是複合音源診斷、隔音設施設計、車道別貢獻分析、熱點治理或工程前後改善評估，陣列方法通常明顯優於單點噪音計。

但也必須精準指出，波束成形與聲學相機並不等於「天然可做法定量測」。第一，這類方法常以相對能量或成像強度呈現結果，未必直接等於法規所要求的受體端聲級。第二，空間解析度受到陣列孔徑、頻率範圍、感測器數量、旁瓣、遮蔽、地面反射與近遠場假設影響。第三，移動車輛場景還要處理都卜勒效應與時間同步問題。正因如此，工程上常把波束成形用作來源定位與工程診斷，再由標準聲

級計負責最終的合規值判讀。這種「量值由聲級計負責，空間歸因由陣列負責」的邏輯，才是目前最穩健、最可防禦的做法。

如果進一步談到「聲源分離」，就不能只停留在幾何定位。因為道路噪音常見的困難，不只是兩個來源位於不同方向，而是它們在同一時間、相近頻帶、相似包絡下重疊，例如兩輛機車同向加速、汽車與重車並行、同一輛車的輪胎噪音與排氣／引擎噪音重疊，或道路與高架橋再輻射同時存在。這時候，光靠方位角還不夠，需要在時域、頻域、時頻域甚至深度特徵空間中把混合信號拆解。Mato-Méndez 等人以雙通道錄音結合盲分離來改善交通噪音自動分類，顯示分離可提升車輛分類與統計資訊提取的效果；這代表聲源分離在道路系統中最實際的角色，往往是作為分類前處理或歸因前處理，而不一定是直接輸出給法規端使用。

就方法論來說，盲源分離（Blind Source Separation）、獨立成分分析（ICA）、非負矩陣分解（NMF）與其後續深度學習版本，本質上都在處理「觀測到的是混合訊號，而來源訊號本身不可直接觀測」的問題。傳統方法的優點是數學假設較清楚，對小型系統較容易解釋；缺點是當道路聲景高度非平穩、來源數量不明、殘響與反射強、或訓練場景與部署場景差異很大時，效能會快速下降。新一代深度學習分離則能用更強的非線性能力處理重疊來源，但代價是需要大量標註或半標註資料，也更容易碰到泛化能力與可解釋性問題。因此，若要把聲源分離作為未來複合性音源治理工具，最佳位置不是立即取代法定測值，而是先進入「輔助量測層」，專責處理混合聲景拆解、來源貢獻排序與後端辨識精度提升。

再往下一層，就是「聲源辨識」。辨識與分離不同，分離重點是把訊號拆開，辨識重點是回答「這是什麼」。在道路場景中，辨識可以有不同粒度。最粗略的是把聲音分成道路交通、鐵路、航空、施工、鳥鳴或人聲等大類；再細一層是車種分類，如機車、小客車、大客車、重型車；更細者則是聲學機制分類，如輪胎—路面、動力總成、煞車、排氣、橋面結構輻射等。Mesaros 等人的聲音事件偵測（SED）教程很清楚地說明，這類系統的核心任務是辨識「什麼事件在何時發生」，所以它天然適合處理道路系統中的事件性噪音、喇叭、急加速、重車通過或特定異常聲事件。[12] 但必須注意，分類或偵測模型輸出的是類別與時間標記，不是天然具有法律效力的環境音量值；它對治理決策極有幫助，對法定裁罰則通常仍需回接至規範化量測與證據鏈。

對多音源道路系統而言，一個特別值得重視的方向，是把「空間分離」與「語意辨識」耦合起來。也就是先用陣列或多感測器取得方向與位置線索，再讓分類器

對某個空間分區或某個分離後音軌做辨識。這比直接對原始混合聲做分類通常更穩健，因為模型不必同時承擔全部重疊聲源的辨識負擔。Murovec 等人提出的陣列式自動環境噪音量測系統，就結合了自動排除非相關噪音事件、主導來源方向辨識與事件分類三個功能，目標是降低人工判讀負擔並提升環境噪音監測可靠性。這對道路系統有很高啟發性：未來真正有價值的不是單一演算法，而是「法定聲級量測+空間定向+事件分類+來源貢獻估算」的一體化架構。

若進一步問「是否可用於法定量測」，最精確的回答應該是這樣。第一，符合 IEC 61672 與相應國內規範的 Class 1 噪音計，當然可直接作為法定量測主體。第二，陣列麥克風與聲學相機若只是輸出熱圖與相對強度，原則上不能直接替代法定 Leq，但可以作為法定量測之輔助模組，特別適合說明主導方向、干擾排除與治理對象。第三，盲源分離與深度學習辨識原則上屬診斷與佐證工具，現階段宜作為「技術輔助」而非「法定主體」。第四，若法規本身已為某種科技執法工具建立明確程序、排除條件與判定鏈，例如我國車輛噪音聲音照相制度，那麼該工具可以在「特定制度」下具有實質執法功能，但這並不代表它已全面取代一般環境音量標準之法定量測架構。這種區分非常重要，因為它同時保留法規穩定性與技術創新空間。

以我國現況來說，《環境音量標準》已明文要求道路環境音量測定使用 Class 1 噪音計，並對測高、測點與測時作出規範；這意味著一般道路環境噪音之法定主體，仍是標準化聲級量測。但在車輛噪音科技執法上，環境部又已透過聲音照相制度建立另一套執法邏輯，包括背景音修正、前後三秒比對排除環境干擾、天候限制，以及在多車同時經過時結合影像資料與噪音時序圖辨識違規改裝情形。這正好說明一個制度設計原則：當科技工具要進入執法端，不是只靠技術本身，而是必須搭配法規授權、證據鏈設計、排除誤判條件與程序正義。對道路系統多音源量測未來若要正式納入制度，也應循這條路徑前進，而不是直接宣稱「某模型辨識準確，所以可直接當法定量測」。

從工程治理的角度再往前一步看，真正完整的道路系統多音源架構，應該至少包含五個連續模組。第一是法定基底模組，即 Class 1 噪音計或等效固定站，用來取得可比較、可稽核的聲級與頻譜。第二是空間量測模組，也就是陣列麥克風、聲學相機或方向估測系統，用來掌握多車道、多高度、多結構面的空間分布。第三是聲源分離模組，用 NMF、ICA、陣列空間濾波或深度學習把重疊來源拆解。第四是辨識模組，以 SED、ASC、CNN、LSTM 或音視融合模型把來源加上語

意標籤。第五則是責任與治理模組，把上述結果重新投影到車種別、時段別、車道別、設施別與受體別的管理決策中，最後仍以法定量測值作為行政與司法上最穩定的落點。這樣的架構既能保住法律穩定，也能真正處理複合音源現場的複雜性。

若把問題聚焦到「可作為未來複合性音源分離工具」這一點，我會建議優先發展三種組合，而不是孤立看某一項技術。第一種組合是「Class 1 噪音計+大型麥克風陣列／聲學相機」，它最適合做道路多車道、高架平面並存、交流道與橋下場景的空間歸因。第二種組合是「陣列前端+深度學習分離+事件辨識」，它最適合高度重疊的都市交通場景，能把空間資訊與語意資訊耦合。第三種組合是「音訊+影像+法規程序」，也就是類似聲音照相或未來更完整的音視同步執法系統，它最適合需要把噪音事件綁到特定車輛、特定時刻與特定證據鏈的行政場景。其中，第一種最成熟，第二種最具研究潛力，第三種最接近執法實務。

至於聲學向量感測器（Acoustic Vector Sensor, AVS），它值得列入「未來工具」而不宜過早列為「現行法定主體」。原因是 AVS 能在單點同時量測聲壓與粒子速度，因此在方向判別、被動監測、低頻感知與交通流監測上有相當潛力。AIP 的研究已指出，AVS 可用於交通量估計與車輛計數，雖然其車速量測精度仍不如多普勒雷達，但它具有不主動發射訊號、較不受電磁干擾、且能進一步分析音訊特徵等優點。換言之，AVS 對於道路多音源的「被動式方向感知」很有前景，但若要進入法定量測，還需要更完整的標準化、校正流程、重現性驗證與法規銜接。

綜合來看，對道路系統多音源空間聲音量測、聲源分離與辨識的完整判斷應是：法定量測的核心仍應由標準化聲級計掌握，因為它直接對應法規的受管制量與程序要求；麥克風陣列與聲學相機最適合回答「從哪裡來、哪裡最強、哪個部位主導」；盲源分離與深度學習最適合處理「混在一起的聲音如何拆開」；聲音事件偵測與分類最適合回答「那是什麼聲音、什麼時候出現」；音視融合最適合回答「是哪一個對象造成、能否形成執法證據鏈」。因此，在制度上最成熟的策略不是選邊站，而是分層整合：用法定量測守住法律穩定，用空間量測強化診斷，用分離與辨識提升治理精度，用特定程序化科技執法處理需要對象綁定的車輛噪音場景。這樣的架構，才最有可能成為未來道路複合性音源分離與治理的可操作版本。

十、面向法定量測與智慧聲景治理的整合展望

從上表可清楚看出，當前具備法定量測地位的技术皆圍繞在積分型噪音計(IEC 61672)、ISO 362/11819 系列標準化試驗，以及 CNOSSOS-EU 計算模型之上 [7, 8, 12]。這些方法的共通點在於其計量可追溯性(Metrological Traceability)、國際互認(International Mutual Recognition)與長期穩定性，而非空間解析度或聲源辨識能力。另一方面，波束成形、Ambisonics、盲訊號分離與深度學習聲源辨識等先進技術雖在空間解析、音源分離與語意辨識上遠勝傳統法定方法，但因其精度隨訓練資料分佈漂移(Domain Shift)、模型不確定性量化困難、缺乏跨實驗室再現性標準等原因，短期內仍屬於研究、診斷與監測輔助之範疇 [43, 46, 54]。

然而，國際聲學界正積極推動將 AI 式方法納入輔助法定量測的過渡框架。CNOSSOS-EU 的最新修訂討論中已出現將神經網路排放模型作為替代經驗公式之提議 [13]。同時，Jagatheesaperumal 等(2023)提出之智慧城市路面品質聲學評估方法若能通過國際實驗室間比對(Interlaboratory Comparison, ILC)，則可望成為 ISO 11819 系列之 AI 擴充版本 [38]。對於未來複合性音源分離與辨識工具之設計，本文建議採取分層融合(Layered Fusion)策略：於最底層保留 IEC 61672 Class 1 可追溯之聲壓量測鏈以確保法定效力 [7]；於中層以 Ambisonics 或高密度 MEMS 陣列擷取空間資訊 [23, 24, 25]；於上層應用多通道 NMF、Conv-TasNet 或 Att-TasNet 等深度學習分離模型獲得個別聲源訊號 [29, 30, 31, 32, 33]；再於頂層以 PANNs、E-PANNs、MAT-SED 或多頭注意力 Transformer 等預訓練大模型進行語意分類與定位 [43, 44, 49, 51]；最終以 SELD 架構整合空間與語意資訊，提供三維聲景的完整解析 [54, 55, 56]。此一架構既可滿足環境噪音指令之策略性噪音地圖法定需求，又可為都市聲景治理、智慧交通管理與噪音源究責提供前所未有的資訊深度，並能透過邊緣 AI 與 TinyML 實現大規模部署 [64, 65, 66]。

#### 十一、研究缺口與未來方向

綜合本文所整理之近五年文獻，可識別出幾個明確之研究缺口值得後續投入。首先，法定量測與 AI 方法之橋接標準尚未建立；儘管 ISO 362-1:2022 已修訂支援新能源車量測，但 AI 聲源辨識結果之不確定度量化與計量追溯仍缺乏統一框架 [18, 21]。其次，電動車新類型聲學特徵資料集嚴重不足；隨著聯合國歐洲經濟委員會(UNECE) R138 聲學車輛警示系統(Acoustic Vehicle Alerting System, AVAS)之強制導入，電動車低速段聲學特徵與識別研究應加速進行 [50]。其三，跨國大規模標註資料集依舊匱乏；SONYC-UST 專注紐約市、UrbanSound8K 為

單一城市錄音，歐亞非各洲都市聲景之多樣性未被充分反映，此影響模型之地理可遷移性 [60, 61]。其四，SELD 於真實交通情境之長時追蹤仍具挑戰；DCASE 挑戰雖已推進至三維 SELD，但多目標連續追蹤(Multi-target Tracking)於都市噪音中之表現尚未被系統性評估 [54, 56]。其五，邊緣 AI 與法定計量之整合部署需要突破；雖有多篇研究證實 MCU 上之聲分類可行，但如何設計同時滿足 IEC 61672 Class 2 計量規範與 AI 辨識需求之混合型感測器，仍為關鍵工程挑戰 [64, 66]。最後，AI 方法與公共衛生證據之閉環整合仍待實現；雖然健康統合分析已明確建立道路交通噪音與心血管、睡眠之劑量-反應關係 [2, 3, 4, 5]，但如何將高解析度聲源辨識結果轉化為疾病風險分層評估，仍是跨學科合作之未竟之業。

## 十二、結論

本文整合了道路交通噪音領域從法定量測、空間聲學量測、聲源分離、聲音事件辨識、自監督學習到邊緣部署之完整技術譜系，並特別納入近五年(2021–2026)之最新研究進展。WHO 與多篇統合分析之證據明確建立了道路交通噪音對人群健康之顯著負擔 [1–5]，而國際標準體系(IEC 61672、ISO 362、ISO 11819、CNOSSOS-EU)至今仍為法定管制之支柱 [6, 7, 8, 12]。然而，傳統單點噪音計已無法滿足智慧城市時代對「誰在噪音、何時噪音、何地噪音」之精細化治理需求，波束成形、Ambisonics、深度學習分離與預訓練大模型之融合路徑已成為不可逆之技術趨勢 [17–67]。展望未來，分層融合架構結合 IEC 61672 可追溯量測、Ambisonics 空間擷取、Conv-TasNet/Att-TasNet 聲源分離、PANNs/E-PANNs 預訓練辨識、SELD 三維事件定位與 TinyML 邊緣部署，將成為智慧聲景治理之核心技術路徑。此一整合性工具鏈不僅為環境噪音工程師、車輛認證實驗室與環境主管機關提供前所未有的分析深度，亦為公共衛生決策、都市設計與交通政策之實證基礎提供可持續精進的技術基座。

## 參考文獻

- [1] World Health Organization Regional Office for Europe. (2018). Environmental noise guidelines for the European region. WHO Press. <https://iris.who.int/handle/10665/279952>
- [2] Smith, M. G., Cordoza, M., & Basner, M. (2022). Environmental noise and effects on sleep: An update to the WHO systematic review and meta-analysis. *Environmental Health Perspectives*, 130(7), 076001. <https://doi.org/10.1289/EHP10197>
- [3] Fu, X., Wang, L., Yuan, L., Hu, H., Li, T., Zhang, J., Ke, Y., & Wang, M. (2023). Long-term exposure to traffic noise and risk of incident cardiovascular diseases: A systematic review and dose-response meta-analysis. *Journal of Urban Health*, 100(5), 1124–1141. <https://doi.org/10.1007/s11524-023-00769-0>
- [4] Chen, X., Liu, M., Zuo, L., Wu, X., Chen, M., Li, X., Wang, T., Wu, S., & Yang, H. (2023). Environmental noise exposure and health outcomes: An umbrella review of systematic reviews and meta-analysis. *European Journal of Public Health*, 33(4), 725–731. <https://doi.org/10.1093/eurpub/ckad044>
- [5] Pershagen, G., Pyko, A., Aasvang, G. M., Babisch, W., Clark, C., Hansell, A. L., Héritier, H., Janssen, S., Koppelman, G. H., Laszlo, H. E., Lercher, P., Selander, J., Sørensen, M., Thacher, J., & Vienneau, D. (2025). Road traffic noise and incident ischemic heart disease, myocardial infarction, and stroke: A systematic review and meta-analysis. *Environmental Epidemiology*, 9(6), e017. <https://doi.org/10.1097/EE9.00000000000017>
- [6] Braun, M. E., Walsh, S. J., Horner, J. L., & Chuter, R. (2013). Noise source characteristics in the ISO 362 vehicle pass-by noise test: Literature review. *Applied Acoustics*, 74(11), 1241–1265. <https://doi.org/10.1016/j.apacoust.2013.04.005>
- [7] International Electrotechnical Commission. (2013). IEC 61672-1:2013 Electroacoustics — Sound level meters — Part 1: Specifications. IEC.
- [8] de León, G., Del Pizzo, L. G., Teti, L., Moro, A., Bianco, F., Fredianelli, L., & Licitra, G. (2020). Evaluation of tyre/road noise and texture interaction on

- rubberised and conventional pavements using CPX and profiling measurements. *Road Materials and Pavement Design*, 21(Suppl. 1), S91–S102. <https://doi.org/10.1080/14680629.2020.1735493>
- [9] Li, D. (2022). Advancement of close-proximity (CPX) measurement methodology for tyre/road noise [Doctoral dissertation, The Hong Kong Polytechnic University]. PolyU Institutional Repository. <https://theses.lib.polyu.edu.hk/handle/200/11676>
- [10] Baumgartner, B., Fuchs, A., & Haider, M. (2023). Correlation of tyre/road noise measurements via machine learning algorithms. In *Proceedings of Forum Acusticum 2023. European Acoustics Association*. <https://dael.euracoustics.org/confs/fa2023/data/articles/001071.pdf>
- [11] Leung, R. C. K., Rai, S., & Li, D. F. (2024). On application of sound power radiation of rolling tyre measured using CPX-based methodology. *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, 270(1), 1856–1865.
- [12] Peeters, B., & van Blokland, G. (2018). Correcting the CNOSSOS-EU road noise emission values. In *Proceedings of Euronoise 2018 (Paper 211)*. European Acoustics Association. [https://www.euronoise2018.eu/docs/papers/211\\_Euronoise2018.pdf](https://www.euronoise2018.eu/docs/papers/211_Euronoise2018.pdf)
- [13] Martinez, P. R., & King, E. A. (2026). Road noise in Europe: The case for transparent and standardised data presentation. *Acta Acustica*, 10(1), 1. <https://doi.org/10.1051/aacus/2026001>
- [14] Umar, I. K., Adamu, M., Mostafa, N., Riaz, M. S., & Ibrahim, Y. E. (2024). The state-of-the-art in the application of artificial intelligence-based models for traffic noise prediction: A bibliographic overview. *Cogent Engineering*, 11(1), 2297508. <https://doi.org/10.1080/23311916.2023.2297508>
- [15] Goncharenko, D. (2025). Application of machine learning to traffic noise prediction: A case study of Tampere and Oulu [Master's thesis, University of Oulu]. Oulu Repository. <https://oulurepo.oulu.fi/handle/10024/57312>
- [16] Kumar, A., Singh, D. K., Amrendra, K., Gautam, S., Kumar, D., & Singh, M. (2025). A machine learning model to uncover the impact of traffic noise

- annoyance in the inhabitants of Ranchi, Jharkhand. In *Recent Advances in Interdisciplinary Studies and Developments (RAISD-25)*. Atlantis Press.
- [17] Michel, U. (2006). History of acoustic beamforming. In *Proceedings of the 1st Berlin Beamforming Conference (BeBeC)*. German Acoustical Society. <https://elib.dlr.de/47021/>
- [18] Czuchaj, M., & Sarradj, E. (2024). Reliable car pass-by source levels from microphone array data. In *Proceedings of the 10th Berlin Beamforming Conference (Paper BeBeC-2024-D08)*. <https://www.bebec.eu/fileadmin/bebec/downloads/bebec-2024/papers/BeBeC-2024-D08.pdf>
- [19] Knappe, F., Becker, V., Puhle, C., Nolte, T., Meyer, S., & Sarradj, E. (2023). Partial sound source estimation with Helmholtz inverse beamforming as a part of pass-by noise virtualization. *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, 268(3), 2453–2462.
- [20] Fredianelli, L., Bernardini, M., Del Pizzo, L. G., Fidecaro, F., & Licitra, G. (2023). Acoustic source localization in ports with different beamforming algorithms. *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, 265(5), 4827–4838.
- [21] Pinel, L., Bouley, S., Purier, A., Tiffeneau, E., & Lepercque, F. (2023). Acoustic imaging and machine learning for sources localization and identification: Application to in situ vehicle pass-by measurements. In *Proceedings of Forum Acusticum 2023*. European Acoustics Association. <https://dael.euracoustics.org/confs/fa2023/data/articles/000162.pdf>
- [22] Kron, T., Meyer, D. J., Koch, T., Ende, C., & Egeler, J. (2025). Measurement-based sound source modelling of moving vehicles for auralizations utilizing acoustic beamforming. In *Proceedings of Forum Acusticum 2025*. <https://dael.euracoustics.org/confs/fa2025/data/articles/000550.pdf>
- [23] Rafaely, B., Tourbabin, V., Habets, E., Ben-Hur, Z., Lee, H., Gamper, H., Arbel, L., Birnie, L., Abhayapala, T., & Samarasinghe, P. (2022). Spatial audio signal processing for binaural reproduction of recorded acoustic scenes — review and challenges. *Acta Acustica*, 6, 47. <https://doi.org/10.1051/aacus/2022040>

- [24] Georgiou, F., Kawai, C., Schäffer, B., & Pieren, R. (2024). Replicating outdoor environments using VR and ambisonics: A methodology for accurate audio-visual recording, processing and reproduction. *Virtual Reality*, 28, Article 124. <https://doi.org/10.1007/s10055-024-01003-1>
- [25] Pecioski, D., Gavriloski, V., Domazetovska, S., Nikolovski, F., Cvetanoski, V., & Mitik, B. (2024). Development of low-cost MEMS microphone array for sound localization in urban environments. *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, 270(7), 4532–4541.
- [26] Chakravarthula, P., D'Souza, J. A., Tseng, E., Bartusiak, J., & Heide, F. (2023). Seeing with sound: Long-range acoustic beamforming for multimodal scene understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 982–991). <https://doi.org/10.1109/CVPR52729.2023.00102>
- [27] Zhang, T., Guo, Z., Fu, Z., Yang, C., Jia, Y., & Liu, J. (2026). A sound-based vehicle position localization dataset and combined filtering strategy. *ACM Transactions on Sensor Networks*. <https://doi.org/10.1145/3799720>
- [28] Saruwatari, H., Kurita, S., Takeda, K., Itakura, F., Nishikawa, T., & Shikano, K. (2003). Blind source separation combining independent component analysis and beamforming. *EURASIP Journal on Advances in Signal Processing*, 2003(11), 569270. <https://doi.org/10.1155/S1110865703305104>
- [29] Ozerov, A., & Févotte, C. (2009). Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(3), 550–563. <https://doi.org/10.1109/TASL.2009.2031510>
- [30] Févotte, C., Vincent, E., & Ozerov, A. (2018). Single-channel audio source separation with NMF: Divergences, constraints and algorithms. In S. Makino (Ed.), *Audio source separation* (pp. 1–24). Springer. [https://doi.org/10.1007/978-3-319-73031-8\\_1](https://doi.org/10.1007/978-3-319-73031-8_1)
- [31] Luo, Y., & Mesgarani, N. (2019). Conv-TasNet: Surpassing ideal time–frequency magnitude masking for speech separation. *IEEE/ACM Transactions on*

Audio, Speech, and Language Processing, 27(8), 1256–1266. <https://doi.org/10.1109/TASLP.2019.2915167>

- [32] Ravenscroft, W., Goetze, S., & Hain, T. (2022). Att-TasNet: Attending to encodings in time-domain audio speech separation of noisy, reverberant speech mixtures. *Frontiers in Signal Processing*, 2, 856968. <https://doi.org/10.3389/frsip.2022.856968>
- [33] Shi, H., Wu, S., Ye, M., & Ma, C. (2024). A speech separation model improved based on Conv-TasNet network. *Journal of Physics: Conference Series*, 2858(1), 012033. <https://doi.org/10.1088/1742-6596/2858/1/012033>
- [34] Ansari, S., Alnajjar, K. A., Khater, T., Mahmoud, S., & Hussain, A. (2023). A robust hybrid neural network architecture for blind source separation of speech signals exploiting deep learning. *IEEE Access*, 11, 100219–100231. <https://doi.org/10.1109/ACCESS.2023.3313972>
- [35] Xie, E. L., & Zhao, C. (2025). Driver speech emotion recognition based on blind source separation and residual neural network. *Journal of Network Intelligence*, 10(1), 203–218.
- [36] Yin, H., Bai, J., Xiao, Y., Wang, H., Zheng, S., Chen, Y., Das, R. K., Liu, C., & Li, H. (2025). Exploring text-queried sound event detection with audio source separation. In *ICASSP 2025 — IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 1–5). IEEE. <https://doi.org/10.1109/ICASSP49660.2025.10889789>
- [37] Ni, J., & Zhou, Z. (2025). Blind source separation and unmanned aerial vehicle classification using CNN with hybrid cross-channel and spatial attention module. *Scientific Reports*, 15, 07946. <https://doi.org/10.1038/s41598-025-07946-y>
- [38] Jagatheesaperumal, S. K., Bibri, S. E., Ganesan, S., & Jeyaraman, P. (2023). Artificial intelligence for road quality assessment in smart cities: A machine learning approach to acoustic data analysis. *Computational Urban Science*, 3, 28. <https://doi.org/10.1007/s43762-023-00104-y>
- [39] Piczak, K. J. (2015). Environmental sound classification with convolutional neural networks. In *2015 IEEE 25th International Workshop on Machine*

- Learning for Signal Processing (MLSP) (pp. 1–6).  
IEEE. <https://doi.org/10.1109/MLSP.2015.7324337>
- [40] Demir, F., Abdullah, D. A., & Sengur, A. (2020). A new deep CNN model for environmental sound classification. *IEEE Access*, 8, 66529–66537. <https://doi.org/10.1109/ACCESS.2020.2984903>
- [41] Mushtaq, Z., & Su, S. F. (2020). Environmental sound classification using a regularized deep convolutional neural network with data augmentation. *Applied Acoustics*, 167, 107389. <https://doi.org/10.1016/j.apacoust.2020.107389>
- [42] Mushtaq, Z., Su, S. F., & Tran, Q. V. (2021). Spectral images based environmental sound classification using CNN with meaningful data augmentation. *Applied Acoustics*, 172, 107581. <https://doi.org/10.1016/j.apacoust.2020.107581>
- [43] Kong, Q., Cao, Y., Iqbal, T., Wang, Y., Wang, W., & Plumbley, M. D. (2020). PANNs: Large-scale pretrained audio neural networks for audio pattern recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28, 2880–2894. <https://doi.org/10.1109/TASLP.2020.3030497>
- [44] Singh, A., Liu, H., & Plumbley, M. D. (2023). E-PANNs: Sound recognition using efficient pre-trained audio neural networks. *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, 268(1), 5951–5960. <https://doi.org/10.48550/arXiv.2305.18665>
- [45] Choudhary, S., Karthik, C. R., Lakshmi, P. S., & Kumar, S. (2022). LEAN: Light and efficient audio classification network. In *2022 IEEE 19th India Council International Conference (INDICON)* (pp. 1–6).  
IEEE. <https://doi.org/10.1109/INDICON56171.2022.10039921>
- [46] Chan, T. K., & Chin, C. S. (2020). A comprehensive review of polyphonic sound event detection. *IEEE Access*, 8, 103339–103373. <https://doi.org/10.1109/ACCESS.2020.2999388>
- [47] Pankajakshan, A., Bear, H. L., & Benetos, E. (2019). Polyphonic sound event and sound activity detection: A multi-task approach. In *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)* (pp. 323–327). IEEE. <https://doi.org/10.1109/WASPAA.2019.8937193>

- [48] Shabbir, A., Cheema, A. N., Ullah, I., Almanjahie, I. M., & Bashir, F. (2024). Smart city traffic management: Acoustic-based vehicle detection using stacking-based ensemble deep learning approach. *IEEE Access*, 12, 35189–35203. <https://doi.org/10.1109/ACCESS.2024.3371243>
- [49] Reza, S., Ferreira, M. C., Machado, J. J. M., & Tavares, J. M. R. S. (2025). Road traffic events monitoring using a multi-head attention mechanism-based Transformer and temporal convolutional networks. *IEEE Transactions on Intelligent Transportation Systems*, 26(9), 12543–12556. <https://doi.org/10.1109/TITS.2025.11079780>
- [50] Hvastja, A., Ćirić, D., Milivojević, M., & Prezelj, J. (2025). Assessing air and noise pollution through acoustic classification of vehicles fuel types using deep learning. *Heliyon*, 11(8), e01811. <https://doi.org/10.1016/j.heliyon.2025.e01811>
- [51] Cai, P., Song, Y., Li, K., Song, H., & McLoughlin, I. (2024). MAT-SED: A masked audio Transformer with masked-reconstruction based pre-training for sound event detection. In *Proceedings of Interspeech 2024* (pp. 1050–1054). <https://doi.org/10.48550/arXiv.2408.08673>
- [52] Alsouda, Y., Pllana, S., & Kurti, A. (2019). IoT-based urban noise identification using machine learning: Performance of SVM, KNN, bagging, and random forest. In *Proceedings of the International Conference on Omni-Layer Intelligent Systems (COINS '19)* (pp. 62–67). ACM. <https://doi.org/10.1145/3312614.3312631>
- [53] Teng, J., Zhang, C., Gong, H., & Liu, C. (2024). Machine learning-based urban noise appropriateness evaluation method and driving factor analysis. *PLOS ONE*, 19(10), e0311571. <https://doi.org/10.1371/journal.pone.0311571>
- [54] Yeow, J. W., Tan, E. L., Bai, J., Peksi, S., & Gan, W. S. (2024). Squeeze-and-excite ResNet-Conformers for sound event localization, detection, and distance estimation for DCASE 2024 challenge. *DCASE 2024 Technical Report*. <https://doi.org/10.48550/arXiv.2407.09021>
- [55] Jiang, Y., Wang, Q., Du, J., Hu, M., Hu, P., Liu, Z., & Lee, C. H. (2024). Exploring audio-visual information fusion for sound event localization and detection in low-resource realistic scenarios. In *2024 IEEE International*

- Conference on Multimedia and Expo (ICME) (pp. 1–6).  
IEEE. <https://doi.org/10.1109/ICME57554.2024.10687782>
- [56] Dong, Y., Wang, Q., Hong, H., Jiang, Y., Hu, M., Du, J., & Lee, C. H. (2025). An experimental study on joint modeling for sound event localization and detection with source distance estimation. In ICASSP 2025 — IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 1–5).  
IEEE. <https://doi.org/10.1109/ICASSP49660.2025.10887663>
- [57] Zhang, S., Zhang, Y., Liao, Y., Pang, K., Wan, Z., & Zhou, S. (2024). Polyphonic sound event localization and detection based on Multiple Attention Fusion ResNet. *Mathematical Biosciences and Engineering*, 21(2), 2190–2208. <https://doi.org/10.3934/mbe.2024096>
- [58] Yang, B., & Li, X. (2024). Self-supervised learning of spatial acoustic representation with cross-channel signal reconstruction and multi-channel Conformer. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 32, 4363–4376. <https://doi.org/10.1109/TASLP.2024.10675425>
- [59] Santos, O., Rosero, K., Masiero, B., Marchi, E., & Iser, B. (2025). Towards large-scale self-supervised learning for spatial audio: Generating first-order ambisonics from monophonic recordings. *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, 272(3), 4512–4521.
- [60] Cartwright, M., Mendez, A. E. M., Cramer, A., Lostanlen, V., Dove, G., Wu, H. H., Salamon, J., Nov, O., & Bello, J. P. (2019). SONYC urban sound tagging (SONYC-UST): A multilabel dataset from an urban acoustic sensor network. In *Proceedings of the DCASE 2019 Workshop* (pp. 35–39). New York University. <http://archive.nyu.edu/handle/2451/60776>
- [61] Bai, J., Chen, J., & Wang, M. (2022). Multimodal urban sound tagging with spatiotemporal context. *IEEE Transactions on Cognitive and Developmental Systems*, 15(2), 555–565. <https://doi.org/10.1109/TCDS.2022.9737309>
- [62] Tailleur, M., Lagrange, M., Aumond, P., & Tourre, V. (2023). Spectral transcoder: Using pretrained urban sound classifiers on undersampled spectral representations. In *Proceedings of the 8th DCASE Workshop* (pp. 176–180). <https://hal.science/hal-04178197>

- [63] Tailleur, M. (2025). Large-scale monitoring of acoustic sources based on data-driven approaches for urban soundscape analysis, visualization, and summarization [Doctoral dissertation, École Centrale de Nantes]. HAL. <https://hal.science/tel-05536184>
- [64] Bellafkir, H., Vogelbacher, M., & Freisleben, B. (2024). Urban sound classification on resource-constrained edge devices. In *Service-Oriented Computing — ICSOC 2024 (Lecture Notes in Computer Science, Vol. 15404, pp. 347–365)*. Springer. [https://doi.org/10.1007/978-981-96-7238-7\\_19](https://doi.org/10.1007/978-981-96-7238-7_19)
- [65] Grau-Haro, J., Ribes-Serrano, R., Cobos, M., & Alonso, J. J. (2025). Comprehensive evaluation of CNN-based audio tagging models on resource-constrained devices. arXiv Preprint. <https://doi.org/10.48550/arXiv.2509.14049>
- [66] Munirathinam, R., & Vitek, S. (2026). Optimizing CNN-GRU hybrid ratios for resource-constrained audio classification: A systematic study from parameter efficiency to MCU deployment. *IEEE Access*, 14, 12543–12562. <https://doi.org/10.1109/ACCESS.2026.11397605>
- [67] Zakaria, A. A., Amr, T., & Ragheb, A. A. (2025). IoT in smart urban planning: A comprehensive review of applications, developments and engineering perspectives. *IEEE Access*, 13, 102457–102478. <https://doi.org/10.1109/ACCESS.2025.11104244>