Spatial separation of sound sources

Bin Dong

THÈSE

SPATIAL SEPARATION OF SOUND SOURCE

présentée devant

l'Institut National des Sciences Appliquées de Lyon

par

Bin DONG

Master de L'Institut de technologie de Harbin

pour obtenir

le GRADE DE DOCTEUR

école doctorale :

Mécanique, énergétique, Génie Civil, Acoustique

Spécialité : **Acoustique**

Thèse préparée au Laboratoire Vibrations Acoustique

soutenue le 14/04/2014 devant la Commission d'examen

jury

Jérôme ANTONI (Professeur)	INSA de Lyon	Directeur de thèse
Mohamed EL BADAOUI (Professeur)	IUT de Roanne	Rapporteur
Jean-Hugh THOMAS (Maître de Conférences)	LAUM	Rapporteur
Charles PEZERAT (Professeur)	LAUM	Examinateur
Quentin LECLERE (Maître de Conférences)	INSA de Lyon	Examinateur

INSA Direction de la Recherche - Ecoles Doctorales – Quinquennal 2011-2015

SIGLE	ECOLE DOCTORALE	NOM ET COORDONNEES DU RESPONSABLE
CHIMIE	**CHIMIE DE LYON** **http://www.edchimie-lyon.fr** Sec :Renée EL MELHEM Bat Blaise Pascal 3e etage Insa : R. GOURDON	M. Jean Marc LANCELIN Université de Lyon – Collège Doctoral Bât ESCPE 43 bd du 11 novembre 1918 69622 VILLEURBANNE Cedex Tél : 04.72.43 13 95 directeur@edchimie-lyon.fr
E.E.A.	**ELECTRONIQUE, ELECTROTECHNIQUE, AUTOMATIQUE** **http://edeea.ec-lyon.fr** Secrétariat : M.C. HAVGOUDOUKIAN eea@ec-lyon.fr	M. Gérard SCORLETTI Ecole Centrale de Lyon 36 avenue Guy de Collongue 69134 ECULLY Tél : 04.72.18 60.97 Fax : 04 78 43 37 17 Gerard.scorletti@ec-lyon.fr
E2M2	**EVOLUTION, ECOSYSTEME, MICROBIOLOGIE, MODELISATION** **http://e2m2.universite-lyon.fr** Insa : H. CHARLES	Mme Gudrun BORNETTE CNRS UMR 5023 LEHNA Université Claude Bernard Lyon 1 Bât Forel 43 bd du 11 novembre 1918 69622 VILLEURBANNE Cédex Tél : 06.07.53.89.13 e2m2@ univ-lyon1.fr
EDISS	**INTERDISCIPLINAIRE SCIENCES-SANTE** **http://www.ediss-lyon.fr** Sec : Insa : M. LAGARDE	Mme Emmanuelle CANET-SOULAS INSERM U1060, CarMeN lab, Univ. Lyon 1 Bâtiment IMBL 11 avenue Jean Capelle INSA de Lyon 696621 Villeurbanne Tél : 04.72.68.49.09 Fax :04 72 68 49 16 Emmanuelle.canet@univ-lyon1.fr
INFOMATHS	**INFORMATIQUE ET MATHEMATIQUES** **http://infomaths.univ-lyon1.fr** Sec :Renée EL MELHEM Bat Blaise Pascal 3e etage infomaths@univ-lyon1.fr	Mme Sylvie CALABRETTO LIRIS – INSA de Lyon Bat Blaise Pascal 7 avenue Jean Capelle 69622 VILLEURBANNE Cedex Tél : 04.72. 43. 80. 46 Fax 04 72 43 16 87 Sylvie.calabretto@insa-lyon.fr
Matériaux	**MATERIAUX DE LYON** **http://ed34.universite-lyon.fr** Secrétariat : M. LABOUNE PM : 71.70 –Fax : 87.12 Bat. Saint Exupéry Ed.materiaux@insa-lyon.fr	M. Jean-Yves BUFFIERE INSA de Lyon MATEIS Bâtiment Saint Exupéry 7 avenue Jean Capelle 69621 VILLEURBANNE Cedex Tél : 04.72.43 83 18 Fax 04 72 43 85 28 Jean-yves.buffiere@insa-lyon.fr
MEGA	**MECANIQUE, ENERGETIQUE, GENIE CIVIL, ACOUSTIQUE** **http://mega.universite-lyon.fr** Secrétariat : M. LABOUNE PM : 71.70 –Fax : 87.12 Bat. Saint Exupéry mega@insa-lyon.fr	M. Philippe BOISSE INSA de Lyon Laboratoire LAMCOS Bâtiment Jacquard 25 bis avenue Jean Capelle 69621 VILLEURBANNE Cedex Tél :04.72 .43.71.70 Fax : 04 72 43 72 37 Philippe.boisse@insa-lyon.fr
ScSo	**ScSo*** **http://recherche.univ-lyon2.fr/scso/** Sec : Viviane POLSINELLI Brigitte DUBOIS Insa : J.Y. TOUSSAINT	M. OBADIA Lionel Université Lyon 2 86 rue Pasteur 69365 LYON Cedex 07 Tél : 04.78.77.23.86 Fax : 04.37.28.04.48 Lionel.Obadia@univ-lyon2.fr

*ScSo : Histoire, Géographie, Aménagement, Urbanisme, Archéologie, Science politique, Sociologie, Anthropologie

认认真真读书，
清清白白做人。

—— 谨遵祖训

Work hard and Keep clear.

—— For my family

Contents

List of acronyms iv
List of symbols iii
Abstract i
Résumé iii
Résumé détaillé v
Part I: Theory 1
Chapter 1 Introduction 2
1.1 Objectives 2
1.2 Literature survey on inverse acoustical problem 3
1.3 Literature survey on blind source separation 5
1.4 Sate-of-the-art of source separation applied to acoustics 7
1.5 Current challenges 9
1.6 Organization of the dissertation 11
Chapter 2 Backpropagation and decorrelation of sound sources 13
2.1 The forward and inverse acoustical problems 13
2.2 Backpropagation with optimal spatial basis 14
2.3 Decorrelation of sound sources 20
2.3.1 Statistical formalization of mutual incoherence 20
2.3.2 Decorrelation from eigenvalue decomposition 21
2.4 Virtual sources versus actual sources 22
2.5 Blind separation of sources with SOBI 23
Chapter 3 Blind separation of disjoint sources from spatial decorrelation 26
3.1 When are virtual sources coinciding with actual sources? 26
3.1.1 The general case 26
3.1.2 An important particular case: spatially uncorrelated sources 27
3.2 Separation of spatially disjoint sources 29
3.2.1 Enforcing spatial decorrelation with joint diagonalization 30
3.2.2 Joint approximate diagonalization 30
3.3 How does backpropagation affect the spatial disjoint at low frequency? 31
Chapter 4 Blind separation of compact sources from the principle of least spatial complexity 33
4.1 Least spatial variance 33
4.2 Least spatial entropy 35
4.2.1 Introduction to spatial entropy 35
4.2.2 The criterion of least spatial entropy 36
4.2.3 Reorder the primary elements with spatial entropy 37
4.3 Optimization Strategy 39
4.3.1 Optimization in the Stiefel manifold 40
4.3.2 Optimal step size 42
Chapter 5 Determination of the number of sources 44
5.1 Literature survey 44
5.2 AIC and MDL 46

5.3 The eigenvalue spectrum 47
5.4 The cumulative power distrubtion 48
5.5 The entropic L-curve 48
Part II: Experiments 51
Chapter 6 Experiment validation 52
6.1 Laboratory experiments 52
6.1.1 Experimental apparatus 52
6.1.2 Separation from single statistical decorrelation 54
6.1.3 Separation from joint statistical and spatial decorrelation 63
6.1.4 Separation from the principle of least spatial complexity 75
6.1.5 Ranking eigen elements according to increasing spatial entropy 87
6.2 Numerical experiments: simulation of the separation of multipoles 94
6.3 An industrial example 98
6.3.1 Determination of the number of sources 98
6.3.2. Source separation from spatial criteria 100
6.3.3 Ranking eigen elements according to increasing spatial entropy 105
Chapter 7 Parametric analysis 108
7.1 Robustness to the number of sources 108
7.1.1 Laboratory experiments in the near-field 108
7.1.2 Laboratory experiments in the far-field 110
7.1.3 Industrial example 112
7.2 Optimal distance for backpropagation 116
7.2.1 Laboratory experiments in the near-field 116
7.2.2 Laboratory experiments in the far-field 119
7.2.3 An industrial example 121
7.3 Size of the aperture function 124
Conclusions 127
Acknowledgement 130
References 132
Appendices 142
Appendix A 143
Bayesian regularization 143
Appendix B 145
Optimization of CLSV in Stiefiel manifold 145
Appendix C 148
Optimization of CLSE in Stiefiel manifold 148

List of acronyms

AIC: Akaike Information Criterion

BSS: Blind Source Separation

CG: Conjugate Gradient

CLSE: Criterion of Least Spatial Entropy

CLSV: Criterion of Least Spatial Variance

CLT: Central Limit Theorem

CS: Compressed Sensing

DFT: Discrete Fourier Transform

ESM: Equivalent Source Method

EVD: EigenValue Decomposition

HELS: Helmholtz's Equation Least-Squares

IBEM: Inverse Boundary Element Method

ICA: Independent Component Analysis

JAD: Joint Approximate Diagonalization

JADE: Joint Approximate Diagonalization of Eigen-matrices

MDL: Minimum Description Length

NAH: Near-field Acoustical Holography

PCA: Principal Component Analysis

PDF: Probability Density Function

SCA: Sparse Component Analysis

SD: Steepest Descent or Statistical Decorrelation

2SD: Statistical and Spatial Decorrelation

SNR: Signal-to-Noise Ratio

SOBI: Second Order Blind Identification

SONAH: Statistically Optimized NAH

SOS: Second-Order Statistics

STFT: Short Time Fourier Transform

SVD: Singular Value Decomposition

List of symbols

Bold capital letter: matrix variable

Bold lowercase letter: vector variable

Italic letter: scalar variable

X: measurements from sensors

A: mixing matrix

*: mixing operator – product or convolution product

S: source signals

N_s: number of sound sources

s_i: i-th sound source

i: index of sound sources

$\sum_{i=1}^{N_s} \bullet$: sum of variables from i=1 to i=N_s

M: number of microphones

p: sound pressure

p_i : sound pressure from the i-th sound source

$\mathbf{r}_m$: position vector of the m-th microphone

m: index of microphones

ω: angular frequency

ϖ : index of snapshot

r: continuous position vector in source domain Γ

G: Green function of medium

Γ: source domain

ε: measurement noise

N: number of surface elements $\Delta\Gamma$

$d\Gamma$: continuous surface element

$\Delta\Gamma$: discrete surface element

l: index of snapshot surface element $\Delta\Gamma$

$\mathbf{r}_l$: discrete position vector in source domain Γ

Ψ_i: spatial mode of the i-th sound source

α_i: latent variable of the i-th sound source

$\mathbb{E}_{\varpi}\{\bullet\}$: expectation operator over snapshots

δ_{ij} : Kronecker delta

K: order of the expansion

k: index of the order

Φ_k: k-th spatial function

a_{ki}: coefficient of the k-th spatial function Φ_k assigned to the i-th spatial mode Ψ_i

$\mathbf{c}$: vector of coefficients of the sound field

$[\bullet]^+$: pseudo-inverse operator

$\hat{\mathbf{c}}$: estimate of the coefficient vector $\mathbf{c}$

$\hat{}$: symbol of an estimated variable

^{H}: Hermitian transpose operation

λ^2: regularization parameter

$\mathbf{n}$: estimation and/or measurement noise

$|\bullet|$: absolute value

$\pi_i(\mathbf{r})$: i-th source distribution

$\mathbf{e}_i$: i-th column of the identity matrix $\mathbf{I}$

^{T}: transpose operation

$\mathbf{C}_{\hat{\mathbf{c}}\hat{\mathbf{c}}^H}$: correlation matrix of the estimated coefficient vector $\hat{\mathbf{c}}$

σ_n^2 : variance of elements in the noise vector $\mathbf{n}$

$\mathbf{D}$: diagonal matrix made of the singular values of matrix $\mathbf{A}$

$\mathbf{U}$: unitary matrix corresponding to the left factor in the singular value decomposition of matrix $\mathbf{A}$

$\mathbf{V}$: unitary matrix corresponding to the right factor in the singular value decomposition of matrix $\mathbf{A}$

d_i: i-th diagonal element of matrix **D**

$\mathbf{u}_i$: i-th column of matrix **U**

$\mathbf{v}_i$: i-th column of matrix **V**

ω_0: center frequency of a frequency band

z: whitened version of coefficient vector **c**

W: whitening matrix

ν : frequency lag

J: number of frequency lags or pure imaginary number $J = \sqrt{-1}$

j: index of the frequency lag

$\text{off}(\bullet)$: sum of squares of all non-diagonal elements of the cost function

Σ: spatial variance

$\boldsymbol{\Sigma}$: aperture function

P: normalized intensity of a sound source

π_i : i-th source power

$\mathbf{r}_{0i}$: spatial centroid of the i-th source distribution

$Tr\{\bullet\}$: trace operator

log: logarithm in base 10

H: spatial entropy

ln: natural logarithm

$\bar{\mathbf{s}}$: virtual source

$\bar{\boldsymbol{\alpha}}$: virtual latent source

$\{H_{[m]}\}$: set of spatial entropies of virtual sources

$H_{[m]}$: m-th smallest spatial entropy

$\bar{\mathbf{D}}$: reranked matrix **D**

$\bar{\mathbf{U}}$: reranked matrix **U**

$\mathbf{E}_k$: Euclidean gradient

$\frac{\partial H(\mathbf{V}_k)}{\partial \mathbf{V}^H}$: partial derivative of the cost function H with respective to matrix $\mathbf{V}$

$\mathbf{R}_k$: Riemannian gradient

$\mathbf{T}_k$: tangent direction along a geodesic

γ_k : weighting factor

$\langle \mathbf{R}_k, \mathbf{R}_k \rangle_{\mathbf{I}}$: inner product of Riemannian gradient $\mathbf{R}_k$

exp: exponential

μ: step size

$\Re\{\bullet\}$: real part

$\boldsymbol{\Theta}$: geodestic curve

$\boldsymbol{\Lambda}$: semi-positive diagonal matrix

λ_1^2 : first diagonal element of matrix $\boldsymbol{\Lambda}$

$\boldsymbol{\Pi}$: complex permutation matrix

$\mathbf{M}$: matrix of spatial components $\Delta M(\mathbf{r}_k, \mathbf{r}_l)$

$\Delta M(\mathbf{r}_k, \mathbf{r}_l)$: spatial element of a given domain

ϕ : complex phase

$\mathbf{C}^{[k]}(\mathbf{V})$: k-th weighted spatial correlation matrix

$C_{ij}^{[k]}$: (i, j)-th element of $\mathbf{C}^{[k]}(\mathbf{V})$

N_ϖ : total number of snapshots ϖ

$\prod_{i=\hat{N}_s+1}^{M} \bullet$: product of variables from $i=\hat{N}_s+1$ to $i=M$

L: Kullback-Leibler divergence (generally noted as D_{KL})

$|\mathbf{A}|$: determinant of matrix $\mathbf{A}$

$\hat{\mathbf{C}}_{\hat{\mathbf{c}}\hat{\mathbf{c}}^H}^{-1}$: inverse of correlation matrix $\hat{\mathbf{C}}_{\hat{\mathbf{c}}\hat{\mathbf{c}}^H}$

Z: distance between the microphone array and the loudspeaker membranes

D: spacing between the centers of two adjacent loudspeaker membranes

f: working frequency (in Hz)

R: radius of the applied aperture function

ρ_i : spatial correlation coefficient between the i-th sound source and its reference

$\rho_i(\omega)$: spatial correlation spectrum

s_i: i-th actual source

*: complex conjugate

Abstract

Blind source separation is a promising technique for the identification, localization, and ranking of sound sources.

The aim of this dissertation is to offer methods for separating incoherent sound sources which may overlap in both the space and frequency domains by exploiting spatial information. This is found of interest in acoustical applications involving the identification and ranking of sound sources stemming from different physical origins. The fundamental principle of all proposed methods proceeds in two steps, the first one being reminiscent to source reconstruction (e.g. as in near-field acoustical holography) and the second one to blind source separation. Specifically, the source mixture is first expanded into a linear combination of spatial basis functions whose coefficients are set by backpropagating the pressures measured by an array of microphones to the source domain. This leads to a formulation similar, but no identical, to blind source separation. In the second step, these coefficients are blindly separated into uncorrelated latent variables, assigned to incoherent “virtual sources”. These are shown to be defined up to an arbitrary rotation. A unique set of sound sources is finally recovered by searching for that rotation (conjugate gradient descent in the Stiefel manifold of unitary matrices) which minimizes some spatial criteria, such as spatial variance, spatial entropy, or spatial orthogonality. This results in the proposal of three separation criteria coined “least spatial variance”, “least spatial entropy”, and “spatial decorrelation”, respectively. Meanwhile, the condition under which classical decorrelation (principal component analysis) can solve the problem is deduced in a rigorous way.

The same concept of spatial entropy, which is central to the dissertation, is also exploited in defining a new criterion, the entropic L-curve, dedicated to determining the number of active sound sources on the source domain of interest. The idea consists in considering the number of sources that achieves the best compromise

between a low spatial entropy (as expected from compact sources) and a low statistical entropy (as expected from a low residual error).

The proposed methodology is validated on both laboratory experiments and numerical data, and illustrated on an industrial example concerned with the ranking of sound sources on the topside of a Diesel engine. The methodology can also correctly separate very small sources whose amplitudes are 40 dB lower than the strongest sources. At the same time, the robustness to the estimated number of active sources, to the range distance between the source domain of interest and the array of microphones, and to the size of aperture function is demonstrated with success.

Keyword: inverse problem, backpropagation, blind source separation, principle of least spatial entropy, principle of least spatial variance, spatial decorrelation, joint approximate diagonalization

Résumé

La séparation aveugle de sources est une technique prometteuse pour l'identification, la localisation, et la classification des sources sonores.

L'objectif de cette thèse est de proposer des méthodes pour séparer des sources sonores incohérentes qui peuvent se chevaucher à la fois dans les domaines spatial et fréquentiel par l'exploitation de l'information spatiale. De telles méthodes sont d'intérêt dans les applications acoustiques nécessitant l'identification et la classification des sources sonores ayant des origines physiques différentes. Le principe fondamental de toutes les méthodes proposées se décrit en deux étapes, la première étant relative à la reconstruction du champ source (comme par exemple à l'aide de l'holographie acoustique de champ proche) et la seconde à la séparation aveugle de sources. Spécifiquement, l'ensemble complexe des sources est d'abord décomposé en une combinaison linéaire de fonctions de base spatiales dont les coefficients sont définis en rétropropageant les pressions mesurées par un réseau de microphones sur le domaine source. Cela conduit à une formulation similaire, mais pas identique, à la séparation aveugle de sources. Dans la seconde étape, ces coefficients sont séparés en variables latentes décorrélées, affectées à des “sources virtuelles” incohérentes. Il est montré que ces dernières sont définies par une rotation arbitraire. Un ensemble unique de sources sonores est finalement résolu par la recherche de la rotation (par gradient conjugué dans la variété Stiefel des matrices unitaires) qui minimise certains critères spatiaux, tels que la variance spatiale, l'entropie spatiale, ou l'orthogonalité spatiale. Il en résulte la proposition de trois critères de séparation à savoir la “moindre variance spatiale”, la “moindre entropie spatiale”, et la “décorrélation spatiale”, respectivement. De plus, la condition sous laquelle la décorrélation classique (analyse en composantes principales) peut résoudre le problème est établit de une manière rigoureuse.

Le même concept d'entropie spatiale, qui est au cœur de cette thèse, est également

exploité dans la définition d'un nouveau critère, la courbe en L entropique, qui permet de déterminer le nombre de sources sonores actives sur le domaine source d'intérêt. L'idée consiste à considérer le nombre de sources qui réalise le meilleur compromis entre une faible entropie spatiale (comme prévu à partir de sources compactes) et une faible entropie statistique (comme prévu à partir d'une faible erreur résiduelle).

La méthode proposée est validée à la fois sur des expériences de laboratoire et des données numériques et illustrée par un exemple industriel concernant la classification des sources sonores sur la face supérieure d'un moteur Diesel. La méthodologie peut également séparer, de façon très précise, des sources dont les amplitudes sont de 40 dB inférieur aux sources les plus fortes. Aussi, la robustesse vis-à-vis de l'estimation du nombre de sources actives, de la distance entre le domaine source d'intérêt et le réseau de microphones, ainsi que de la taille de la fonction d'ouverture est démontrée avec succès.

MOTS-CLES: problème inverse, rétropropagation, séparation aveugle de sources, principe de moindre entropie spatiale, principe de moindre variance spatiale, décorrélation spatiale, diagonalisation conjointe approchée

Résumé détaillé

La localisation et l'identification de sources sonores font partie des problèmes fondamentaux de l'ingénierie vibro-acoustique. Le besoin de techniques nouvelles pour localiser, identifier et classer les sources sonores traduit des exigences de plus en plus strictes en termes de qualité acoustique, en particulier dans l'industrie du transport. Parmi les différentes approches dédiées à ces objectifs, une solution populaire consiste à reconstruire les sources sonores (par exemple la pression pariétale ou la composante normale de la vitesse des particules) indirectement à partir de mesures faites à distance, telles que celles réalisées par un réseau de microphones. L'approche a l'avantage d'être non intrusive et assez universelle dans son principe. Pour ces raisons, cette technique conue sous le nom de l'imagerie acoustique est largement utilisée dans l'industrie.

Une limitation importante de l'imagerie acoustique classique est qu'elle identifie difficilement les sources sonores qui se chevauchent tant dans les domaines spatial et fréquentiel, même lorqu'elles proviennent d'origines physiques distinctes. Un exemple typique est donné dans l'industrie automobile par les sources sonores (par exemple, la combustion, l'injection, le système d'échappement, etc.) rayonnées par un moteur Diesel .

L'objectif de cette thèse est de combler cette lacune en proposant une solution qui permette de séparer des sources *incohérentes* à partir de mesures à distance; l'incohérence mutuelle est considérée ici comme une propriété exclusive des sources ayant des origines physiques distinctes. Après avoir parcouru la littérature existante sur la séparation aveugle de sources (Blind Source Separation, BSS) en acoustique, il semble qu'au moins trois questions doivent être considérées dans un effort de développement de méthodes autonomes ayant un potentiel réel dans les applications industrielles:

1) la première question est de savoir comment séparer aveuglément des sources

sonores incohérentes dans le domaine fréquentiel sans références et sans avoir recours à des hypothèses particulières telles que la non-gaussiennatié;

2) la deuxième question, qui est consécutive à la première, est de savoir si il y a une condition spécifique sous laquelle la décorrélation standard peut résoudre directement le problème de BSS sans aucune information supplémentaire;

3) la troisième question, lée aux deux premières, est comment déterminer de façon fiable le nombre de sources sonores incidentes sur un réseau de microphones avec une précision acceptable, comme condition préalable à la BSS.

Le manuscrit de thèse est divisée en deux parties: théorie et expériences. La partie I se compose des Chapitres 1-5 qui sont principalement concentrés sur la théorie de la séparation spatiale des sources sonores incohérentes. La partie II est constitué les Chapitres 6-7, et se concentre sur la validation des algorithmes proposés dans la partie I par des expériences de laboratoire, par la simulation numérique, et par un exemple industriel – séparation aveugle de sources sonores rayonnées par un moteur Diesel. En outre, l'analyse des paramètres expérimentaux par rapport aux algorithmes proposés est étudiée en fin de manusscrit. L'introduction concrète de chaque chapitre est présentée respectivement comme suit:

Le Chapitre 1 évoque d'abord l'objectif de recherche de la présente thèse, c'est-à-dire la séparation aveugle de sources sonores incohérentes qui se chevauchent dans les deux domaines spatial et fréquentiel. Ensuite, la littérature sur le problème acoustique inverse est arpentée en détail. Ia discussion se concentre principalement sur la formation de voies (beamforming) et l'holographie acoustique de champ proche (Near-Field Acoustical Holography, NAH), qui sont les deux méthodes les plus couramment appliquées dans l'imagerie acoustique. Dans le paragraphe suivant, l'état de l'art sur la séparation aveugle de sources est abordé en trois points: le modèle, le critère de séparation aveugle de sources, et l'algorithme d'optimisation. Les applications de la séparation de sources en acoustique, principalement supervisée et aveugle, sont étudiés en détail, respectivement. Sur la base de la littérature existante, les défis actuels sont identifiés au paragraphe 1.5. L'organisation du reste du manuscrit est comme suit.

La théorie de base sur “la séparation spatiale des sources sonores” est introduite au Chapitre 2, qui se structure en deux parties: la rétropropagation et la décorrélation statistique des sources sonores. Dans un premier temps, les problèmes directs et inverses en acoustique sont examinés de façon concise. Le modèle mathématique général de la rétropropagation est établi comme une introduction au problème inverse. Pour mettre en évidence les performances de cette dernière méthodologie, un ensemble de bases spatiales optimales, déduites de la théorie bayésienne, est adopté pour rétropropager les ondes sonores à partir des mesures réalisées par un réseau de capteurs (par exemple, microphones) sur le domaine source d'intérêt. La décorrélation statistique des sources sonores est réalisée par décomposition en valeurs propres de la matrice de corrélation des coefficients se rapportant à la base spatiale optimale. Les sources virtuelles issues de la décorrélation statistique sont ensuite comparés aux sources réelles. On montre qu'étant donné une matrice de rotation manquante, les sources virtuelles ne correspondent pas aux sources réelles dans le cas général. Pour séparer les sources sonores par utilisation de la seule information statistique basée sur les sources virtuelles, une méthode aveugle classique au second ordre (Second Order Blind Identification, SOBI) est adoptée pour séparer aveuglément les sources sonores réelles à la fin du Chapitre 2.

En se basant sur la théorie du Chapitre 2, les conditions nécessaires pour que les sources virtuelles coïncident bien avec les sources réelles sont étudiées du point de vue mathématique au début du Chapitre 3. Il est montré que celles-ci sont liées à la notion d'orthogonalité spatiale des sources, un cas particulier intéressant étant celui de sources spatialement disjointes. Cependant, cette propriété peut être éventuellement perdue lors de la rétropropagation en champ lointain. Pour améliorer les résultats de la séparation de sources spatialement disjointes, un algorithme de décorrélation conjointe statistique et spatiale et proposé sur la base d'une diagonalisation conjointe approchée (Joint Approximate Diagonalization, JAD) ou de la méthode du gradient conjugué (Conjugate Gradient, CG) en fin du Chapitre 3. Avec l'aide de l'information spatiale, les sources réelles peuvent ainsi être séparées avec succès sur la plupart des fréquences de travail, même basses.

La méthode proposée dans le Chapitre 3 ne peut cependant pas traiter les scénarios difficiles, par exemple avec grand chevauchement spatial ou avec un mélange de sources de niveaux très différents. Pour résoudre les problèmes mentionnés ci-dessus, deux critères de séparation spatiale sont introduits dans le Chapitre 4, à savoir le critère de la moindre variance spatiale (Criteria of Least Spatial Variance, CLSV) et le critère de la moindre entropie spatiale (Criterion of Least Spatial Entropy, CLSE), sur la base d'une hypothèse intuitive – la compacité spatiale. Ce critère est intéressant d'un point de vue pratique, car l'étendue spatiale des sources réelles est souvent réduite en raison de leurs origines physiques et coïncide par ailleurs bien avec le sens de l'audition humaine.

Inspiré par le concept important de *variance*, le CLSV est présenté dans la section 4.1, en rapport avec la distribution spatiale des sources d'intérêt. La variance spatiale mesure la distribution spatiale des sources sonores autour de leurs centroïdes, où le centroïde peut être considéré comme l'espérance spatiale de la source sonore. Il est évident que la distribution de la source la plus compacte correspond à la variance spatiale minimale. Dans un tel critère de séparation, la matrice de rotation manquant peut finalement être déterminée par minimisation de la variance spatiale totale des sources sonores. Le CLSV a l'avantage dêtre simple, mais il n'est pas adapté à la séparation des sources sonores compactes mais multimodales.

Par conséquent, le CLSE est proposé pour résoudre les problèmes dans lesquels le CLSV est limité ou défaillant. Avant de présenter le CLSE, le concept critique d'*entropie spatiale*, inspirée par l'entropie de Shannon, est introduit au début de la Section 4.2. L'entropie spatiale correspondant aux sources réelles est un minimum parmi les valeurs de toutes les sources suggérées. C'est-à-dire que la redondance spatiale des sources candidates est supérieure ou égale à celle des sources réelles. Ainsi, l'unicité de la solution des sources réelles est garantie par la minimisation de l'entropie spatiale. En appliquant une optimisation similaire à celle de CLSV, chaque source réelle est évaluée en fonction de la matrice de rotation optimale correspondant au minimum de l'entropie spatiale.

Pour plus de simplicité, le CLSV et le CLSE sont dénommés par le principe de

moindre complexité spatiale ou de compacité spatiale. Un organigramme de l'exécution du principe de moindre complexité spatiale est représenté à la Fig. 1.

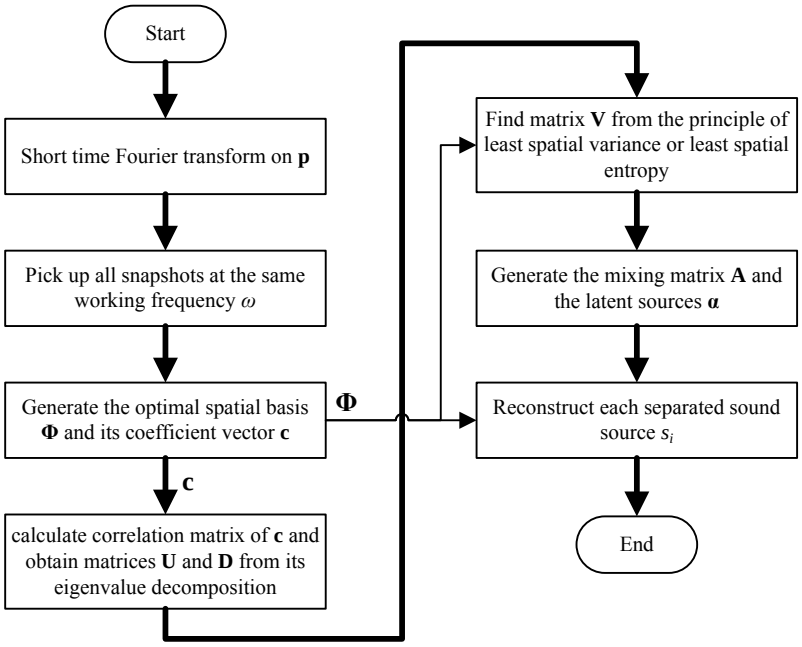

Fig. 1 Organigramme du principe de moindre complexité spatiale.

Lorsque certaines sources sonores sont beaucoup plus petites que les autres, une version améliorée du principe de moindre complexité spatiale est proposée. Les éléments propres sont classés en fonction de leur entropie spatiale croissante au lieu de leurs niveaux comme réalisé de manière standard dans le traitement du signal des réseaux de capteurs. Ainsi, les très petites sources peuvent être reconstituées avec succès au moyen de la matrice latente obtenue et les sources virtuelles associées.

Les fonctions de coûts découlant du principe de moindre complexité spatiale (c'est-à-dire CLSV et CLSE) font face au même problème: comment optimiser les fonctions de coût sous une forte contrainte – la matrice unitaire. Une solution possible est de transformer le problème d'optimisation avec contrainte dans l'espace Euclidien en un problème non-constraint dans la variété Stiefel engendré par les matrices unitaires. L'optimisation des fonctions de coût dans la variété Stiefel est exécutée par un algorithme du gradient conjugué avec un pas de taille optimal. Pour rechercher ce

dernier, deux méthodes, l'approximation polynomiale et la DFT (Transformée de Fourier Discrète). Du point de vue des performances, la méthode DFT est recommandée.

Commun aux autres problèmes de BSS, le nombre de sources sonores doit être déterminé correctement à l'avance dans notre scénario. Un nombre sous-estimé de sources impliquera que quelques sources manqueront dans la séparation; par contre une surestimation du nombre des sources entrainera un calcul trop lourd et long. Avant d'introduire les méthodes pour répondre à cette question, la littérature sur l'état de l'art de la déterminationr du nombre de sources dans le traitement de signal et des antennes de capteurs est parcourue de manière concise. En se basant sur la littérature, quatre type de méthodes sont introduites.

Les méthodes issues de la théorie de l'information impliquent principalement deux critères: le critères d'information d'Akaike (Akaike Information Criterion, AIC) et la longueur de description minimale (Minimum Description Length, MDL). Ces deux critères d'information sont connus pour surestimer le nombre de sources. Une autre approche, purement énergétique, analyse le spectre des valeurs propres et sa puissance cumulée. Les deux méthodes sont confrontés à la même restriction sur le rapport signal sur bruit (Signal-to-Noise Ratio, SNR) qui doit être suffisamment élevé pour qu'il y ait un écart distinct entre l'espace signal et l'espace bruit. Enfin, le même concept d'entropie spatiale, qui est au cœur de la thèse, est exploité pour définir un nouveau critère, la courbe en L entropique, dédié à la détermination du nombre de sources sonores actives dans le domaine source d'intérêt. L'idée consiste à considérer le nombre de sources qui réalise le meilleur compromis entre une faible entropie spatiale (comme prévu à partir de sources compactes) et une faible entropie statistique (comme prévu à partir d'une faible erreur résiduelle).

Les algorithmes proposés dans les Chapitres 3-5 sont illustrés dans le Chapitre 6 sur des expériences de laboratoire, des simulations numériques, et un exemple industriel, successivement. Les performances de tous les algorithmes sont discutées en détail. Pour commencer, le dispositif d'expériences de laboratoire est brièvement introduit. Les Figs. 2-3 illustrent le dispositif expérimental et le schéma de procédé

expérimental, respectivement.

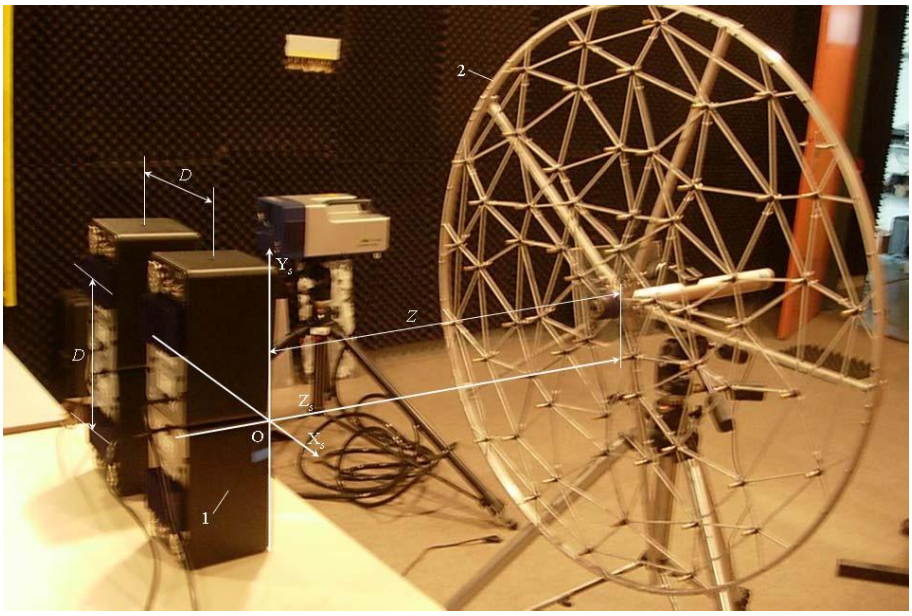

Fig. 2 Le dispositif d'expérimental pour valider l'algorithmes proposés (1: les haut-parleurs, 2: l'antenne de microphones de la roue de coupe).

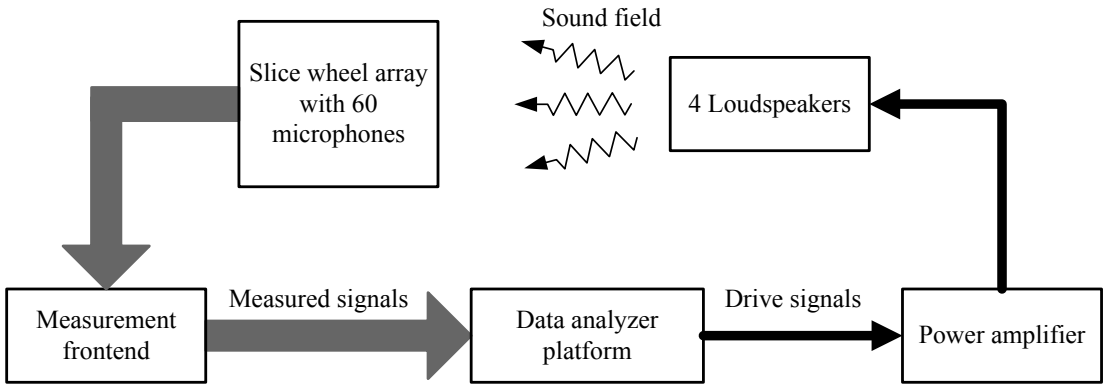

Fig. 3 Schéma du processus expérimental.

Avant de présenter les résultats de la séparation, le nombre de sources actives est déterminé par les quatre méthodes introduites dans le Chapitre 5. Selon l'estimation de ce nombre, les résultats de la séparation dans les expériences de laboratoire sont présentées d'abord pour la décorrélation statistique en champs proche et lointain. Les résultats de séparation impliquent non seulement les sources sonores séparées à la fréquence de travail spécifique, mais aussi les spectres de vitesse quadratique et les

spectres de corrélation spatiale dans toute la bande de fréquence disponible. La Fig. 4 illustre les sources séparées par décorrélation statistique en champ proche. Comme prévu dans le Chapitre 3, la décorrélation statistique fonctionne bien dans le cas où la condition d'orthogonalité spatiale est satisfaite.

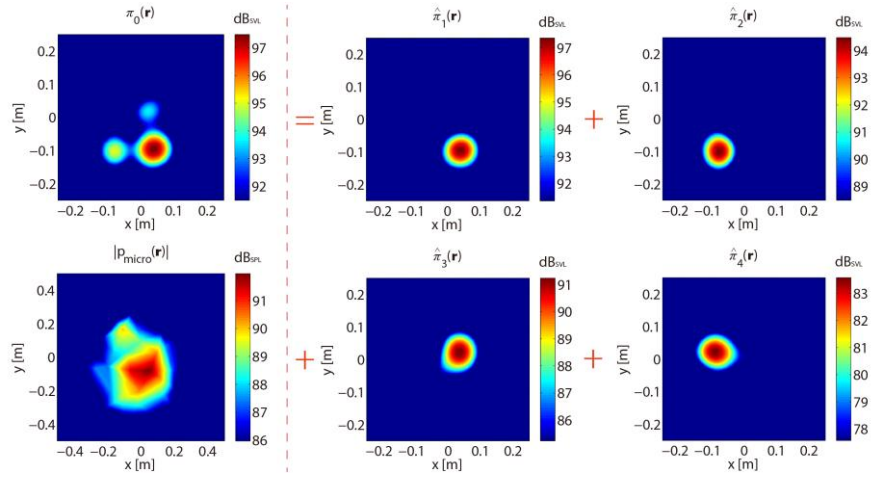

Fig. 4 Sources séparées “virtuelles”, $\hat{\pi}_i(\mathbf{r})$, i = 1,…,4, à partir de la décorrélation statistique, en champ proche à 1800 Hz avec une fonction d'ouverture R = 35 cm..

Pour séparer aveuglément les sources sonores disjointes dans l'espace, le nouveau critère spatial dénommé “décorrélation spatiale” est ensuite appliqué à la fois en champs proche et lointain avec les algorithmes JAD et CG. Les résultats de la séparation (à la fois en champs proche et lointain) sont apparemment améliorés par rapport à la décorrélation statistique seule. Il est à noter qu'il n'y a pas de différence significative entre les résultats issus de JAD et de CG sur les résultats de la séparation. Les résultats sont représentés sur les Figs. 5-6, respectivement.

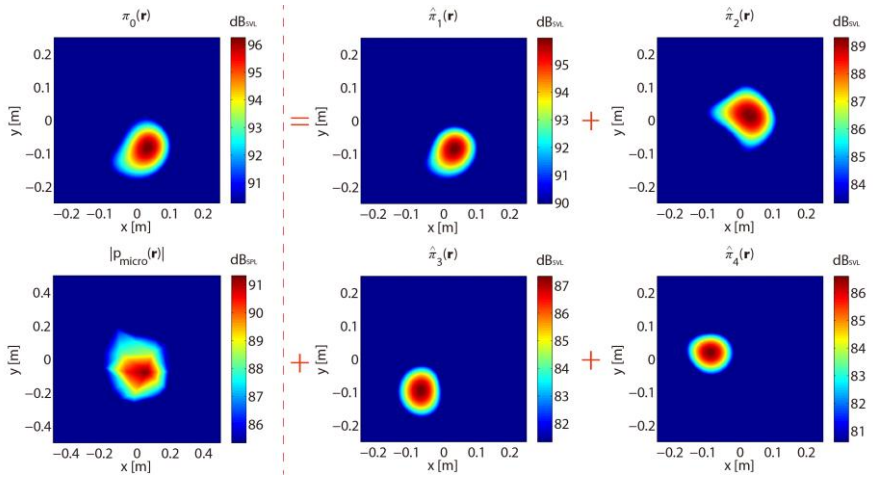

Fig. 5 Sources séparées, $\hat{\pi}_i(\mathbf{r})$, $i = 1,\ldots,4$, à partir de la décorrélation statistique et spatiale avec diagonalisation conjointe approchée, en champ proche à 833 Hz..

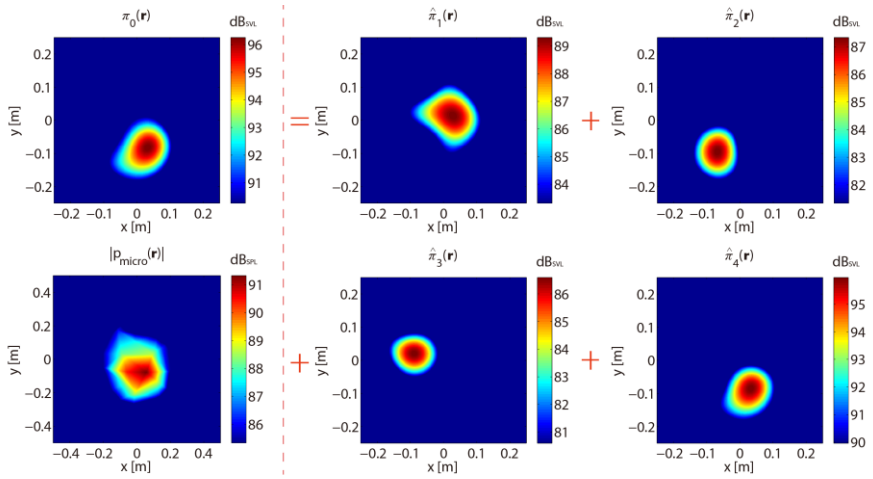

Fig. 6 Sources séparées, $\hat{\pi}_i(\mathbf{r})$, $i = 1,\ldots,4$, à partir de la décorrélation statistique et spatiale avec la méthode de gradient conjugué, en champ proche à 833 Hz..

Le principe de moindre complexité spatiale est alors appliqué sur les expériences de laboratoire dans le paragraphe suivant. Les deux critères spatiaux, CLSV et CLSE, sont validés par ces expériences à la fois en champs proche et lointain. Les sources séparées en champ proche sont représentés sur les Figs. 7-8, correspondant à CLSV et CLSE, respectivement. Le CLSE est recommandé parmi tous les algorithmes proposés en fonction de sa bonne performance.

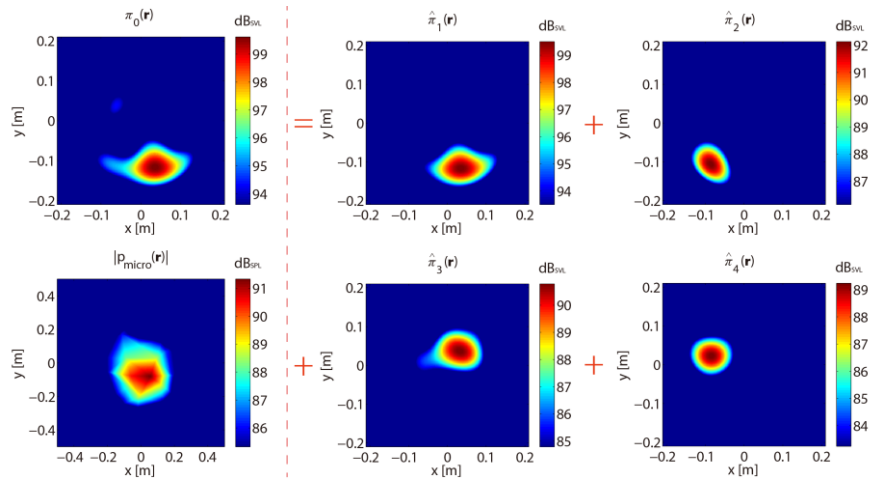

Fig. 7 Sources séparées, $\hat{\pi}_i(\mathbf{r},\omega)$, $i = 1,\ldots,4$, à partir du critère de moindre variance spatiale, en champ proche à 833 Hz..

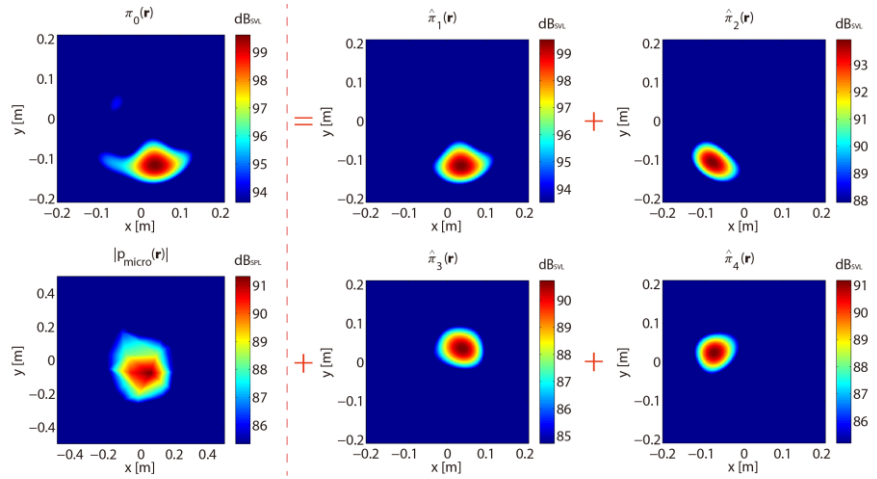

Fig. 8 Sources séparées, $\hat{\pi}_i(\mathbf{r},\omega)$, $i = 1,\ldots,4$, à partir du critère de moindre entropie spatiale, en champ proche à 833 Hz.

Dans le dernier paragraphe de la Section 6.1, une très petite source est séparée avec succès en classant les éléments propres selon l'entropie spatiale croissante des sources virtuelles. Les résultats de la séparation en champ proche sont illustrés dans la Fig. 9 où une source séparée est de 40 dB plus petite que les autres sources.

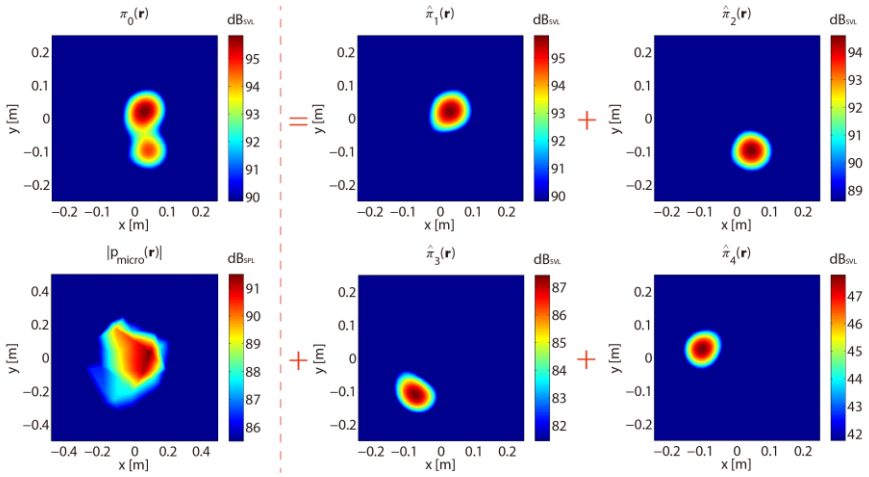

Fig. 9 Sources séparées, $\hat{\pi}_i(\mathbf{r},\omega)$, $i = 1,\ldots,4$, à partir du critère de moindre entropie spatiale, en champ proche à1456 Hz.

Dans la Section 6.2, des expériences numériques sont mises en œuvre pour valider les algorithmes proposés. Ici, CLSE est appliqué pour séparer des multipôles sonores. Le principe de moindre complexité spatiale arrive encore à séparer les multipôles avec succès, même s'ils se chevauchent fortement les uns aux autres dans l'espace.

Un exemple industriel, la séparation des sources actives rayonnée par la face supérieure d'un moteur Diesel, est étudiée à partir des critères spatiaux de décorrélation statistique et spatiale, de moindre variance spatiale, et de moindre entropie spatiale, respectivement. Les sources séparées à partir de CLSE sont montrées dans la Fig. 10. Même les petites sources produites par le moteur Diesel sont séparées avec succès grâce au classement des éléments propres selon l'entropie spatiale.

Le Chapitre 7 concerne l'analyse paramétrique des algorithmes proposés. Le CLSE est l'algorithme choisi ici pour faire l'analyse paramétrique en fonction de ses bonnes performances dans de nombeux contextes. Tout d'abord, la robustesse au nombre de sources est étudiée par des expériences de laboratoire à la fois en champs proche et lointain et sur l'exemple industriel. Les résultats sont globalement stables.

Deuxièmement, la distance optimale de la rétropropagation est étudiée par des expériences de laboratoire et sur l'exemple industriel. La distance optimale de la

rétropropagation est proche de la distance réelle. Le CLSE se trouve robuste par rapport à la distance de la rétropropagation, c'est-à-dire fonctionnant globalement bien non seulement pour des sous-estimées et sur-estimées.

Troisièmement, la taille de la fonction d'ouverture de la rétropropagation est brièvement inspectée sur l'exemple industriel. Les algorithmes proposés peuvent séparer correctement les sources actives qui se situent dans la fonction d'ouverture de la rétropropagation quelle que soit sa taille.

Enfin, les conclusions, les points de vue et quatre points saillants sont discutés en fin de manuscrit.

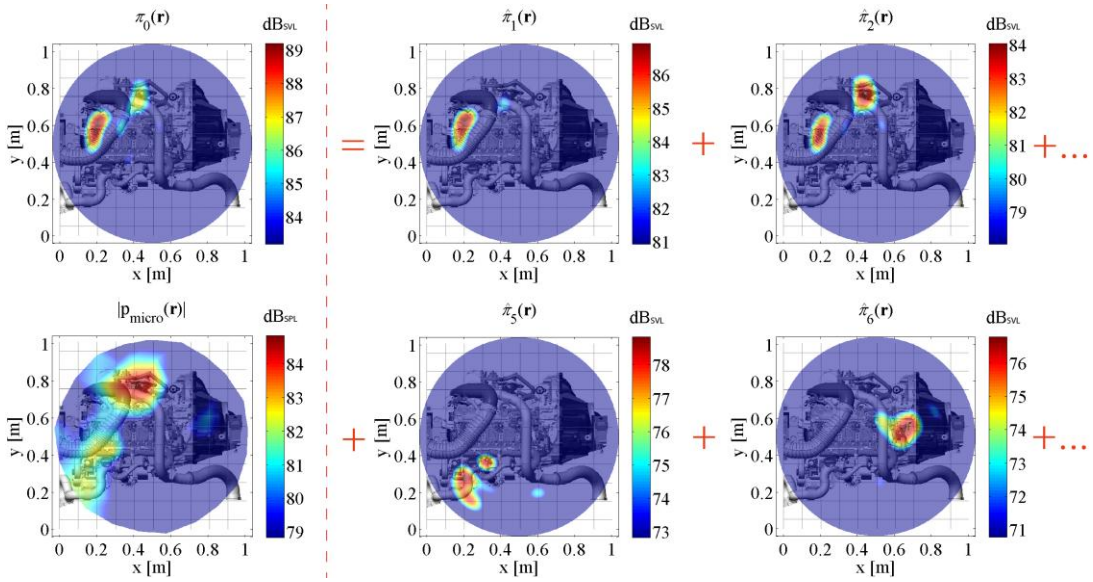

Fig. 10 Sources séparées, $\hat{\pi}_i(\mathbf{r},\omega)$, i = 1,2,5,6, à partir du critère de moindre entropie spatiale (moteur Diesel à 1250 Hz).

Part I: Theory

Chapter 1 Introduction

1.1 Objectives

Localization and identification of sound sources are fundamental issues in sound and vibration engineering [1]. The need for new techniques to localize, identify, and rank sound sources has been recently boosted by stricter requirements in terms of acoustical quality, especially in the transportation industry. Among the various approaches dedicated to these objectives [2]-[4], one popular solution is to reconstruct sound sources (e.g. parietal pressure or normal component of particle velocity) indirectly from contactless measurements, such as returned by an array of microphones. This has the definite advantage of being non-intrusive as well as being rather universal in principle. Acoustic imaging is widely applied in the industry.

A few examples are listed in Fig. 1.1. Figure 1.1(a) displays the sound field distribution on the top side of a flying airliner, where the strong sound sources are marked in red and the low ones in blue. Apparently, the primary sources origin from the entrance of the four aircraft engines and from the head of the plane. The ranking of the sound sources radiated from a wind turbine is conveniently identified from the acoustic image as shown in Fig. 1.1(b). In Fig. 1.1(c), a large source is located on the bottom side of a sewing machine, which mainly supports the machine. Fig. 1.1(d) depicts the spatial distribution of sound sources which are generated by the wheels of a train rolling on rails. The last panel of Fig. 1.1 shows sound sources located on the front side of an automobile engine, which is of interest to survey the performance of the engine.

One significant limitation of classical acoustic imaging is that it can hardly identify sound sources which overlap in both the spatial and frequency domains, even though they may stem from distinct physical origins. A typical example is given in the automotive industry, by the sound sources (e.g. combustion, injection, exhaust system, and etc.) radiated from a Diesel engine [7]-[8].

The aim of the present dissertation is to fill this gap by proposing solutions that

separate *incoherent* sources from remote measurements – mutual incoherence is considered here as an exclusive property of sources originating from distinct physical origins. Before introducing the state-of-the-art of source separation in acoustics, the inverse acoustical problem is surveyed first in the next subsection.

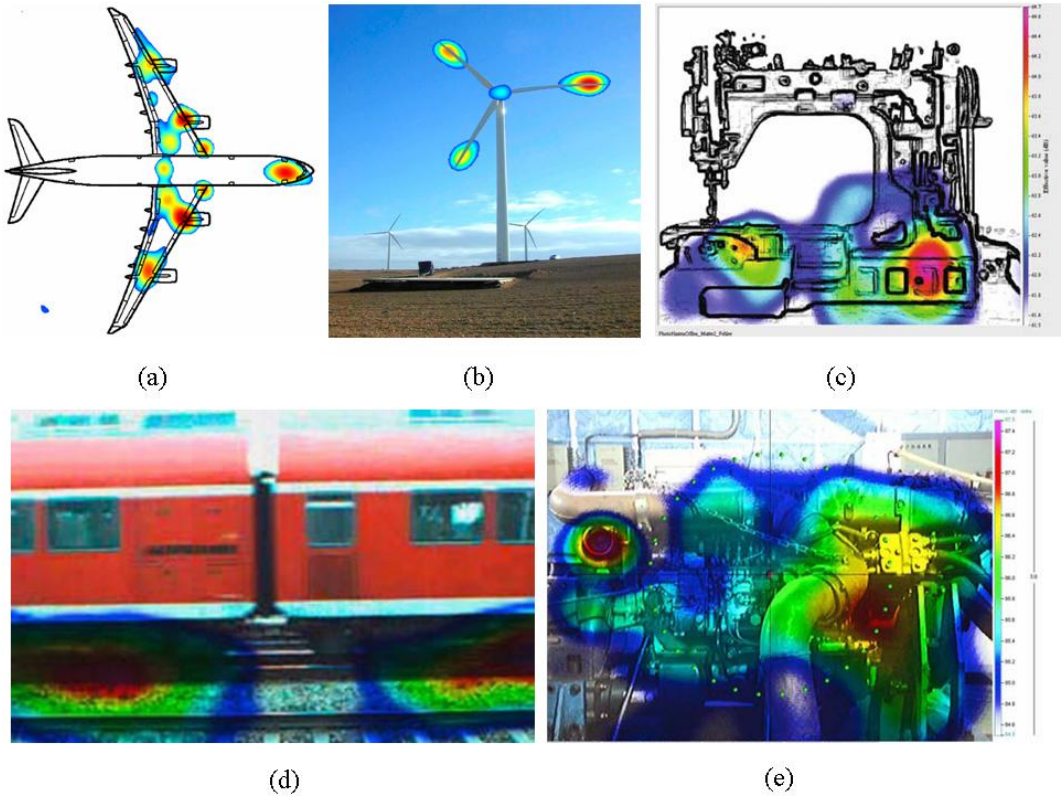

Fig. 1.1 Examples of reconstructing sound sources with an array of microphones: (a) the top side of airliner [5], (b) the blades and rotor of a wind turbine [5], (c) the bottom side of a sewing machine [6], (d) the connection part between rails and wheels of a train [6], and (e) the front side of an engine [6].

1.2 Literature survey on inverse acoustical problem

In essence, the technique of acoustic imaging is an inverse problem. How to solve this kind of problem is out of the scope of the dissertation. Here only the development of acoustic imaging is reviewed from the historical point of view in short.

Beamforming, one of the standard methods for source localization, is undoubtedly

the first proposed technique of acoustic imaging [9]-[10]. Beamforming has a series of advantages: high spatial resolution at high frequency, suitability to several types of sources (no matter moving or stationary), applicability to large target objects (for instance aircraft and wind turbine) to name just a few. At the same time, it also owns several apparent drawbacks, especially a low spatial resolution at low frequency and a limitation to operate in the far-field. To counter the disadvantages of beamforming, other methods have been developed based on other physical properties of sound waves.

By applying the spatial Fourier transform and combining the evanescent waves in the near-field, outstanding successes have been obtained – in particular in terms of attainable spatial resolution – after the introduction of Near-field Acoustical Holography (NAH) by Maynard, Williams and Lee in the 80's [11]-[12]. Elaborating on the idea of backpropagating the measured acoustical field from the array to the source of interest, several versions and variants of NAH have been proposed over the years, such as Statistically Optimized NAH (SONAH) [13]-[14], Helmholtz's Equation Least-Squares (HELS) [15], the Equivalent Source Method (ESM) [16]-[18], and the Inverse Boundary Element Method (IBEM) [19]-[20] to name just a few. Reviews of these methods may be found in Refs. [21]-[22]. NAH works well at low frequency of interest, because of the acquisition of evanescent waves. As like the two sides of a coin, working in the near-field brings many attractive properties to NAH, but meanwhile it becomes the bottleneck in other scenarii.

One noticeable drawback of NAH is that it needs a large number of microphones to break through the wavelength limitation and then reconstruct the acoustic hologram at high frequencies. Taking into account the cost of sensors and the technical limitation of the hardware, simultaneously applying too many measurements is not without drawback for NAH.

Thank to its smart theory [23]-[25], Compressed Sensing (CS) has the potential to solve the contradiction between working at high frequencies and measuring with fewer microphones. G. Chardon *et al.* have successfully demonstrated the possibility of applying CS to NAH [26].

In fact, the inverse acoustic problem can be considered as an *interpolation-extrapolation* problem, when taking into account the process from measurements to the source domain of interest. In the interpolation step, the measurements from an array of microphones are expanded into a set of basis functions. For instance, the measurements are interpolated by typically plane wave functions in SONAH, by typically spherical wave functions in HELS, by monopoles/dipoles in ESM, and by generalized spherical wave functions in IBEM. Appropriate regularization techniques are employed to depress unstable results in the extrapolation step, e.g. Tikhonov regularized solution is considered in SONAH [13], which primarily bases on the type of basis functions in the interpolation step.

A question is naturally proposed: is there an optimal one among all possible candidates of interpolation basis functions, when the topology of a source domain, the geometry of an array of microphones, and any possible *a priori* information on the source distribution in space are available in advance? The answer might be positive. The optimal basis functions for the interpolation have been found by J. Antoni with success under the Bayesian framework [27]. The main work of the dissertation is fully bases on the backpropagation with the optimal basis functions.

For the discussion of inverse acoustical problem in detail, one tutorial [28] and two PhD theses [29]-[30] are recommended here.

1.3 Literature survey on blind source separation

Blind source separation (BSS), a hot research topic in the signal and image processing community [31]-[44], originated at the beginning of the 80's [45].The literature on BSS is extremely vast (it has been growing exponentially since the early 90's). Obviously, it is not necessary and not practical to survey all the literatures concerning BSS. Hereafter the main content of the present section concentrates on the introduction of the framework of BSS from three parts: model, criterion and algorithm.

As declared by P. Comon in Ref. [44], the 'blind' in BSS does not mean absolute

blindness, but implies to make use of *a priori* information as little as possible. In general, the *a priori* information is implicitly involved in the model of BSS. According to the physical nature of the problem of interest, the BSS model is categorized into *linear* versus *nonlinear*, *noisy* versus *noiseless*, *instantaneous* versus *convolutive*, working in the *time* domain versus the *frequency* domain, *underdetermined* versus *overdetermined* in terms of the numbers of sources and sensors… For convenient purpose, the BSS model is usually formulated in the matrix format,

$$\mathbf{X} = \mathbf{A} * \mathbf{S}, \tag{1.1}$$

where, **X** denotes the matrix of measurements from sensors, **A** is the mixing matrix, '*' stands for the mixing operation – for instance, *linear* or *nonlinear*, *product* or *convolution*, and **S** represents the matrix of sources. Obviously, without any complementary information, Eq. (1.1) is impossible to solve precisely, because it has infinite solutions for the *underdetermined* case and an optimal least-squares solution, might far from the actual one, for the *overdetermined* case. Note that ignoring the scale-permutation ambiguity is the specific unspoken rule in BSS. To improve the abstract of the mathematic model, S. Amari and A. Cichocki combine block diagram with mathematic formulas to model the BSS problem from the neural network point of view ([46], [43]).

The basic idea behind all criteria of BSS in the signal and image processing community is to minimize the statistical information provided by sources. The criteria could be roughly divided into two groups according to their statistical property: *Gaussian* and *non-Gaussian* sources. To separate non-Gaussian sources, their higher (normally more than 2)-order statistics are usually involved into the criterion. Since the theory of Independent Component Analysis (ICA) has been built by P. Comon [31], it has become, undoubtedly, the first choice for the blind separation of non-Gaussian sources. To solve the problem of mixed Gaussian sources, the corresponding criteria have to resort to the Second-Order Statistics (SOS), as the higher-order statistics of Gaussian sources are all null.

To optimize the corresponding criteria and solve the modeled BSS problem, a large

number of algorithms have been successively proposed since the birth of BSS. Three classic algorithms, Joint Approximate Diagonalization of Eigen-matrices (JADE), Second Order Blind Identification (SOBI), and Sparse Component Analysis (SCA), are underlined in short here. JADE firstly forms a set of 4th-order cumulants of the whitened measurements and then jointly diagonalizes them to separate the sources [47]. It is based on the two rigorous assumptions of *statistical independence* and *non-Gaussianity* of the sources. SOBI applies the joint approximate diagonalization to a group of whitened covariance matrices of the measurements in order to separate Gaussian sources [48].Interestingly, sources can be considered to be *sparse* (more generally, approximately sparse) in many applications. With the help of such an assumption, they can be separated with SCA [49] even if the assumption of statistical independence is lost. Noteworthily, SCA has immeasurable potential to BSS, as more and more powerful algorithms from CS have been proposed to deal with sparse signals and images [50]-[51].

For the complete history on the development of BSS, the reader is referred to Chapter 1 of "Handbook of Blind Source Separation: Independent Component Analysis and Application" edited by P. Comon and C. Jutten [44].

1.4 State-of-the-art of source separation applied to acoustics

In this dissertation, the research interest is limited to the scenario where all sound sources are located on one side of an array of microphones. If interested to other cases, the readers are referred to E. F. Grande's work [29].

A first possibility towards the separation of sound sources is the "supervised separation" which makes use of a set of available references on the sources of interest in order to separate their contributions at the measurement points. These methods have attracted considerable attention since the late 70's, after the advent of dual channel analyzers [52]. They are, in theory, able to separate overlapping sources (in time, frequency, and space), yet they crucially rely on the availability of pure references, perfectly correlated with the sources of interest and totally uncorrelated

with other sources. Early applications of supervised source separation to NAH are reported in Refs. [53]-[56]. A recent application to the Diesel engine example quoted above is found in Ref. [8]. An important practical limitation of the supervised separation is the difficulty of mounting reference sensors close enough to the sources of interest which, in general, may not be known *a priori*. In Ref. [57], Tomlinson investigated the possibility of obtaining "virtual" sources from the Principal Component Analysis (PCA) or partial coherence analysis of a set of non-ideal reference transducers. However, as early recognized by S. M. Price and R. J. Bernhard [58], there is no guarantee that a virtual source should endow a physical essence. References [59]-[62] tackle the issue from a different perspective, where "numerical" references are obtained from the reconstructed source field after applying NAH in a first step. The method has its elegance and instrumental simplicity, yet it relies on the assumption that references can be assigned to spatially non-overlapping sources in the reconstruction domain and, therefore, can hardly solve the issue raised in the present dissertation.

Since striving for pure enough references is the main bottleneck of supervised source separation, one may wonder whether references are really necessary to achieve separation after all. Relaxing the need for reference is the objective of BSS which has been surveyed shortly in Section 1.3. BSS relies on the paradigm that a mixture of sources can be separated provided that sources bear different enough statistical properties. In a sense, sources should have disjoint support sets in some statistical space (of which trivial examples are the frequency and time domains).

Although the literature on BSS is extremely vast (until June 2009, 22,000 scientific papers concerning BSS have been published in Engineering, Computer Science, and Mathematics according to Google Scholar [44]), few reports have been given to acoustical applications (except for speech processing which, in many aspects, is closer to communication signal processing than to the issue addressed in the present dissertation) and, as far as the authors know, none to NAH. A few examples of separation of sound signals are given in Refs. [63]-[64], although they are quite far from the objectives of the present dissertation. In Ref. [63], Zhong *et al.* apply a

standard BSS algorithm (i.e. SOBI) which forces decorrelation of separated sources at several time-lags); one limitation of the approach is that it assumes an instantaneous mixture of sources, whereas a convolutive mixture should be used as soon as acoustical propagation is taken into account. In Ref. [64], Teramoto and M. Khan propose a BSS method for acoustical signals, where the convolution mixture is turned into an instantaneous mixture by making use of the Euler's equation. The requirement for simultaneous measurements of the sound pressure and of the particle velocity in the three Cartesian coordinates is a strong practical limitation of the method; in addition, it is limited to the separation of directions of arrival in the far-field. A rather advanced BSS scheme is proposed by Aichner *et al.* in Ref. [65], which can separate broadband acoustical sources in reverberant and noisy environments such as in a car compartment. Separation is achieved by forcing the decorrelation of the sources at several time-lags based on a convolutive model. Moreover, a large literature exists on the separation of audio signals which, at first glance, may appear somewhat relation to the objectives of the present dissertation; yet it is not reviewed here since these methods put a high demand on the specific waveform of audio signals (e.g. strong amplitude and frequency modulation) [66]-[68], and thus are likely to fail in any industrial application where sound sources do not necessarily exhibit such features.

One reference that makes explicit account of the acoustical specificities is [69], wherein the authors devised a method based on the Bayesian framework to separate broadband sound sources in the frequency and space domains. However, the method may not be yet ready for industrial applications because of its high calculation burden plus the requirement for some critical parameters to be tuned by the user.

1.5 Current challenges

Reasons why standard BSS algorithms – such as developed in communication engineering – can hardly be applied to separate acoustical sources, are the following.

First, most BSS algorithms require the strong assumption of statistically independent sources, which is much more restrictive than their mutual

incoherence/decorrelation[1] (independence implies decorrelation, but the reverse does not hold true in general); related to this is the necessary condition is that no more than one source in the mixture can be Gaussian [[31], [34], [36], [42]-[44]]. Unfortunately, this is rapidly invalidated when acoustical signals are transformed into the Fourier domain – a common practice in acoustics and the preliminary step in NAH – since the Discrete Fourier Transform (DFT) makes the data quickly converge to a complex Gaussian distribution according to the Central Limit Theorem (CLT) [70]. This precludes the use of ICA, including some popular algorithms such as JADE [47]. Although second-order BSS (e.g. see [[48], [71]]) could separate Gaussian sources by forcing their decorrelation at several time-lags in the time domain (as in aforementioned Refs. [63] and [65]), it can be shown that it does not apply in the frequency domain for *stationary* sound sources (because stationary sources have no correlation in the frequency domain, independently of being separated or not).

A second limitation is that a "source", in acoustics, is endowed with a spatial distribution that is actually inexistent in BSS, where the concept is essentially a statistical quantity that is not necessarily being assigned to a physical essence.

As a consequence, a third limitation is that, in acoustics, spatial propagation from the source to the measurements must be taken into account, typically by means of a Fredholm integral of the first kind (e.g. a "convolution" integral over space), in addition to convolution over time. Related to this the fact is that separation of sound sources is a severely ill-posed problem [72]-[76], especially due to the existence of evanescent waves that do not reach the array [77]. To the authors' knowledge, the modeling of acoustical propagation has rarely been addressed in the BSS literature – except in the very particular case where the measurements are taken in the far-field.

Another usual condition in BSS is that the number of sources is not greater than the number of sensors, or strictly less in the presence of additive noise. Concerning this point, the acoustical context may present an advantage since the habit is to use quite dense arrays, where the number of microphones is likely to overrun the number of

[1] The terms "decorrelation" and "incoherence" are synonymous, the first one being employed in statistics and the second one in acoustics; a formal definition is given in Eq. (2.14).

active sources to be separated.

Finally, a last limitation is that the vast majority of the BSS algorithms require the number of sources to be known exactly. In many acoustical applications, this quantity is actually an unknown parameter that is to be determined from the data beforehand; unfortunately, this turns out to be an extremely difficult problem with no universal solution.

After surveying the existing literature on BSS in acoustics, it appears that at least three issues are to be answered in an endeavor to develop standalone methods with actual potential in industrial applications:

1) the first issue is how to blindly separate incoherent sound sources in the frequency domain without references and without resorting to special assumptions such as non-Gaussianity;

2) the second issue, following the first one, is whether there is a specific condition under which standard decorrelation can directly solve the BSS problem without any additional information;

3) the third issue, interleaved with the first two ones, is how to reliably determine the number of sound sources impinging on an array of microphones with acceptable accuracy, as a prerequisite to BSS.

1.6 Organization of the dissertation

The rest of the dissertation is devoted to proposing solutions to these three issues. It is divided into two parts: Theory and Experiments. The first part involves Chapters 1-5 and the second part Chapters 6-7. The organization of the content is introduced as follows.

Chapter 2 first introduces the direct and inverse acoustical problems. It then formulizes a mathematical model for the BSS problem that specifically accounts for acoustical propagation without restriction to the far-field. Finally, a classic method for BSS – SOBI, is presented to deal with the proposed problem in short.

The main content of Chapter 3 is to answer the second issue. First of all, the

condition under which the virtual sources returned by standard statistical decorrelation coincide with the actual ones is rigorously demonstrated from the mathematical point of view. Afterwards, it presents an original criterion, named spatial decorreltion, to blindly separate sound sources which are spatially disjoint.

Chapter 4 introduces two original criteria that are apt to return unique solutions to the blind separation of incoherent sound sources with compact distributions. At the same time, the order of uncorrelated latent variables is rearranged first in terms of the increasing spatial entropies of virtual sources, and then the reranked virtual source are applied to separate very small sources whose amplitude can be 40 dB lower than the other ones. Lastly, the chapter describes how to optimize the cost functions derived from the two aforementioned original criteria, in the Stiefel manifold, in a concise way.

In Chapter 5, the literature on the determination of the number of sources is first surveyed in short. Then four methods are discussed to determine the number of active sources from the recorded data. Three classic methods, the information theoretical criteria, the eigenvalue spectrum, and the cumulative power distribution, are presented first. An original criterion, coined the *entropic L-curve*, is then proposed at the end of Chapter 5.

Chapter 6 demonstrates the effectiveness of the proposed methodology on laboratory experiments, on a numerical simulation, and on an industrial example. Based on the analysis of the separation results, the criterion of least spatial entropy is recommended in the end.

In Chapter 7, the robustness of the proposed algorithm with respect to the experimental parameters is discussed in light of the results, from laboratory experiments and from an industrial example. The experimental parameters of interest are the number of sources, the range distance between the source domain of interest and the array of microphones, and the size of the aperture function, respectively.

Conclusions and perspectives and four points are drawn at the end of the dissertation.

Chapter 2 Backpropagation and decorrelation of sound sources

In this chapter, the forward and inverse problems in acoustics are introduced first. Based on the introduction of the inverse problem, backpropagation from measurements is presented using a set of optimal spatial bases. Statistical decorrelation of sound sources is conducted in terms of an assumption – 'mutual incoherence' based on backpropagation. The virtual sources and the actual sources are then compared to each other from the points of energy and spatial distribution. Finally, the classic BSS method – SOBI, combining with the spatial distribution of sound sources, are listed in short.

2.1 The forward and inverse acoustical problems

The forward and inverse problems are fundamentals in acoustics. A radiated sound field could be totally determined by a known source distribution, given the medium of propagation. That is the so called the forward problem. On the contrary, the inverse problem is to reconstruct a model to describe the unknown source distribution via a known sound field with the minimum error between the output of the model and the measurements. Figure 2.1 presents an example which identifies a fundamental difference between the forward problem and the inverse problem. The forward problem is to fix the distribution of sound field in space from the known source distribution, and the inverse one is the opposite. There are three sound sources (e.g. the membranes of loudspeakers) in a static homogenous free space, and the sound field generated from the sound sources is recorded by an array of microphones. The forward problem mission is to determine sound signals on all microphones from a known source distribution (e.g. parietal pressure or normal component of particle velocity on the membranes of loudspeakers). Instead, the object of the inverse

problem is much more intractable – i.e. applying measurements on the microphones to estimate the optimal model for the unknown source distribution with the given medium.

In acoustical applications, people meet the inverse problem more frequently. Moreover, the inverse problem is usually ill-posed in mathematical sense, which faces many theoretical limitations, such as the existence of solution, the uniqueness of the solution, and the stability of the solution [78]-[80]. In the past decades, many researchers focused on the inverse problem in acoustics [[81]-[86], [72]-[77]], which is one of the main research topics in source identification and characterization. In this dissertation, the inverse problem is concerned with the backpropagation of measurements sampled by an array of microphones to the sound domain of interest, which will be introduced in the following.

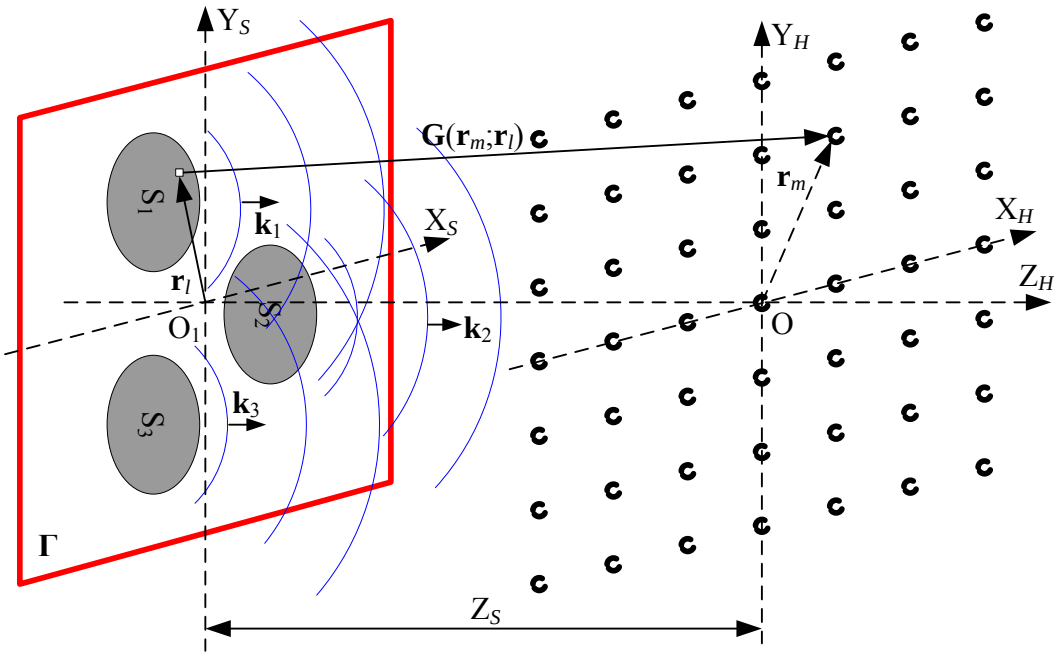

Fig. 2.1 The forward problem and the inverse problem.

2.2 Backpropagation with optimal spatial basis

Rather generally, let us consider the case of N_s incoherent sound sources s_i, $i = 1$,

2, …, N_s, radiating in a static homogenous free space to an array of M microphones ("incoherence" and "decorrelation" are the same to each other in this dissertation); no other sources are assumed in the space between the source domain and the hologram plane, as depicted in Fig. 2.1. Let us denote $p_i(\mathbf{r}_m, \omega; \varpi)$ the sound pressures radiated from the i-th source s_i, i=1,…N_s, at the m-th microphone located at position $\mathbf{r}_m$ (m=1, 2, …, M for a M-microphone array), at angular frequency ω, and indexed by the snapshot number ϖ (in practice, a snapshot corresponds to a short-time block of data and, in theory, to a given realization of a stochastic process; the two concepts coincides when the process is stationary and ergodic, which will be assumed from now on). Figure 2.2 displays how to generate the snapshot of sound pressure signal, $p(\mathbf{r}_m, \omega; \varpi)$, from the m-th channel of the array by using Short Time Fourier Transform (STFT).

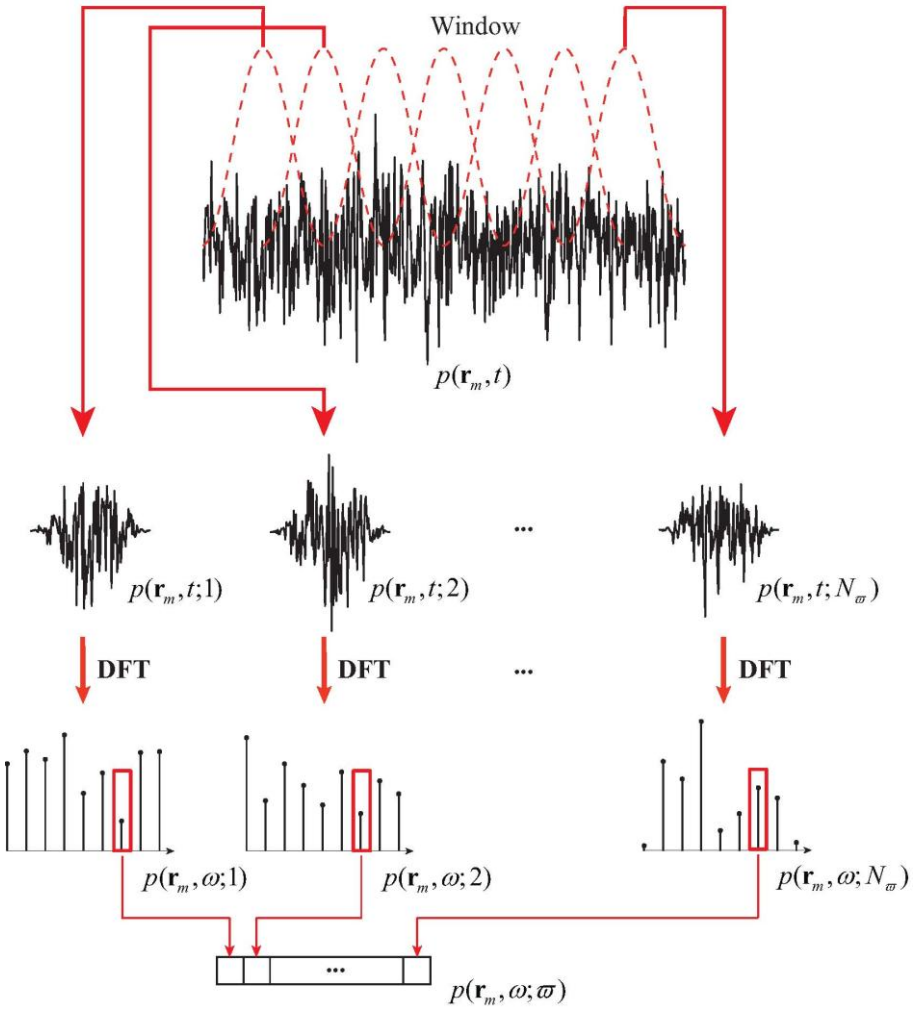

Fig. 2.2 Illustration of the concept of snapshots: $p(\mathbf{r}_m,\omega;\varpi)$ denotes the short-time Fourier transform of the pressure signal from the *m*-th channel of the array, at angular frequency ω, and datum ϖ.

Thus, contribution $p_i(\mathbf{r}_m,\omega;\varpi)$ from the *i*-th source distribution, $s_i(\mathbf{r},\omega;\varpi)$, to the *m*-th microphone is given by the Fredholm integral [87]-[88]

$$p_i(\mathbf{r}_m,\omega;\varpi)=\int_\Gamma s_i(\mathbf{r},\omega;\varpi)G(\mathbf{r}_m,\mathbf{r};\omega)d\Gamma(\mathbf{r}), \tag{2.1}$$

where $G(\mathbf{r}_m,\mathbf{r};\omega)$ stands for the Green function of the medium and Γ for the source domain. In turn, the sound pressure measured by the *m*-th microphone is the sum of the contributions from the N_s sound sources, that is

$$p(\mathbf{r}_m,\omega;\varpi)=\sum_{i=1}^{N_s}p_i(\mathbf{r}_m,\omega;\varpi)+\varepsilon(\mathbf{r}_m,\omega;\varpi), \tag{2.2}$$

where $\varepsilon(\mathbf{r}_m, \omega; \varpi)$ stands for "measurement" noise including instrumental origins as well as modeling errors [89].

Upon substituting Eq. (2.1) into Eq. (2.2) and discretizing the source domain Γ, one obtains the following discrete form

$$p(\mathbf{r}_m, \omega; \varpi) = \sum_{i=1}^{N_s} \sum_{l=1}^{N} s_i(\mathbf{r}_l, \omega; \varpi) G(\mathbf{r}_m, \mathbf{r}_l; \omega) \Delta\Gamma(\mathbf{r}_l) + \varepsilon(\mathbf{r}_m, \omega; \varpi), \tag{2.3}$$

where $\Delta\Gamma(\mathbf{r}_l)$ stands for the surface element on the source domain Γ at position $\mathbf{r}_l$ and N for the total number of such elements. The sound pressure vector corresponding to the whole array can thus be expressed in a matrix form as

$$\mathbf{p}(\omega; \varpi) = \underbrace{\begin{bmatrix} G(\mathbf{r}_1, \mathbf{r}_1; \omega) & G(\mathbf{r}_1, \mathbf{r}_2; \omega) & \cdots & G(\mathbf{r}_1, \mathbf{r}_N; \omega) \\ G(\mathbf{r}_2, \mathbf{r}_1; \omega) & G(\mathbf{r}_2, \mathbf{r}_2; \omega) & \cdots & G(\mathbf{r}_2, \mathbf{r}_N; \omega) \\ \vdots & \vdots & \ddots & \vdots \\ G(\mathbf{r}_M, \mathbf{r}_1; \omega) & G(\mathbf{r}_M, \mathbf{r}_2; \omega) & \cdots & G(\mathbf{r}_M, \mathbf{r}_N; \omega) \end{bmatrix}}_{\mathbf{G}(\omega)} \underbrace{\begin{bmatrix} \Delta\Gamma(\mathbf{r}_1) & & & \\ & \ddots & & \mathbf{0} \\ & & & \\ & \mathbf{0} & & \Delta\Gamma(\mathbf{r}_N) \end{bmatrix}}_{\mathbf{\Gamma}} \underbrace{\begin{bmatrix} s(\mathbf{r}_1, \omega; \varpi) \\ s(\mathbf{r}_2, \omega; \varpi) \\ \vdots \\ s(\mathbf{r}_N, \omega; \varpi) \end{bmatrix}}_{\mathbf{s}(\omega; \varpi)} + \underbrace{\begin{bmatrix} \varepsilon(\mathbf{r}_1, \omega; \varpi) \\ \varepsilon(\mathbf{r}_2, \omega; \varpi) \\ \vdots \\ \varepsilon(\mathbf{r}_M, \omega; \varpi) \end{bmatrix}}_{\boldsymbol{\varepsilon}(\omega; \varpi)} \tag{2.4}$$

where $s(\mathbf{r}_l, \omega; \varpi) = \sum_{i=1}^{N_s} s_i(\mathbf{r}_l, \omega; \varpi)$ is the sum of the N_s sound sources; that is in a more concise form,

$$\mathbf{p}(\omega; \varpi) = \mathbf{G}(\omega)\mathbf{\Gamma}\mathbf{s}(\omega; \varpi) + \boldsymbol{\varepsilon}(\omega; \varpi). \tag{2.5}$$

The ultimate objective is now to separate the sound sources in their domain Γ, i.e. to estimate each s_i from $\mathbf{p}$. The following parameterization is proposed towards this perspective. First, each sound source is modeled as the product of spatial mode Ψ_i denoting the spatial distribution of the source with a variable α_i – referred to in the present work as the "latent variable" – describing the source random amplitude:

$$s_i(\mathbf{r}, \omega; \varpi) = \Psi_i(\mathbf{r}, \omega)\alpha_i(\omega; \varpi). \tag{2.6}$$

Note that the spatial mode Ψ_i depends on space whereas it is independent of the snapshot number, and vice versa for the latent variable α_i; in other words, Eq. (2.6) makes separable variables $\mathbf{r}$ and ϖ. This fully reflects the distinctive property of a sound source to be fully coherent with itself. In addition, the mutual incoherence of

sound sources implies that

$$\mathbb{E}_{\varpi}\left\{\alpha_i(\omega;\varpi)\alpha_j^*(\omega;\varpi)\right\}=\delta_{ij}\,, \tag{2.7}$$

where, $\mathbb{E}_{\varpi}\{\bullet\}$ stands for the expectation operator (ensemble average over snapshots ϖ), superscript '*' for the complex conjugate, and δ_{ij} for the Kronecker delta. Without loss of generality, the power spectra of the latent variables are assumed unitary (i.e. the source magnitude is thus assigned to the spatial mode). Because in most cases it is not known *a pripori*, the spatial mode Ψ_i is further expanded into a series of predefined spatial functions, i.e.

$$\Psi_i(\mathbf{r},\omega)=\sum_{k=1}^{K}a_{ki}(\omega)\Phi_k(\mathbf{r},\omega)\,, \tag{2.8}$$

where K is the order of the expansion and a_{ki} is the coefficient of the k-th spatial function Φ_k assigned to the i-th spatial mode Ψ_i. Many choices are possible for the spatial functions Φ_k; some typical ones are the plane waves, the cylindrical or the spherical harmonics, and layers of monopoles or dipoles. A set of optimal spatial functions with minimum reconstruction error (among all possible choices given a source domain topology and an array geometry) is given in Ref. [27]. This optimal basis will be adopted from now on due to its interesting properties and in order to keep the reconstructed source distribution as precise as possible – however, any other choice would make possible the application of the proposed BSS methods of Chapters 3 and 4. In this particular case, it has been shown that K is equal to the number M of microphones in the array [27].

Combining Eqs. (2.6) and (2.8), the source expansion has the matrix form

$$\mathbf{s}(\omega;\varpi)=\underbrace{\mathbf{\Phi}(\omega)\mathbf{A}(\omega)}_{\mathbf{\Psi}(\omega)}\boldsymbol{\alpha}(\omega;\varpi)=\mathbf{\Phi}(\omega)\mathbf{c}(\omega;\varpi)\,, \tag{2.9}$$

where the (l, k)-th element of matrix $\mathbf{\Phi}$ is $\Phi_k(\mathbf{r}_l,\omega)$, the ($k$, i)-th element of matrix $\mathbf{A}$ is $a_{ki}(\omega)$, and the i-th element of vector $\boldsymbol{\alpha}$ is $\alpha_i(\omega;\varpi)$; this defines

$$\mathbf{c}(\omega;\varpi)=\mathbf{A}(\omega)\boldsymbol{\alpha}(\omega;\varpi) \tag{2.10}$$

as the vector of coefficients of the sound field expanded onto the spatial basis

$\{\Phi_k(\mathbf{r},\omega)\}$. Upon substitution into Eq. (2.5), one arrives at the following propagation model

$$\mathbf{p}(\omega;\varpi)=\mathbf{G}(\omega)\boldsymbol{\Gamma}\boldsymbol{\Phi}(\omega)\mathbf{A}(\omega)\boldsymbol{\alpha}(\omega;\varpi)+\boldsymbol{\varepsilon}(\omega;\varpi)\,. \tag{2.11}$$

Given a set of measured pressures, the Green function of the medium and a spatial basis for expanding the source field, Eq. (2.11) makes it clear that the two unknown quantities are **A** and **α**. Referring back to Eq. (2.10), this resembles the classical BSS problem as stated in signal processing wherein it is conventionally written **x**=**As**, with **A** the "mixing matrix" and **s** the "sources" (see e.g. [90]-[92], [31]-[51]); obvious differences, however, are

i) the premultiplication with the operator **GΓΦ**,

ii) the explicit presence of additive noise, and

iii) the fact that the mixing matrix does not apply directly to the sources but to the latent variables **α** which are connected to the actual physical sources through Eq. (2.9).

Instead of trying to solve Eq. (2.11) directly for **A** and **α**, a somewhat simpler – but suboptimal – strategy is proposed here that consists in uncoupling the problems of reconstruction and separation. This proceeds in two steps. First an estimate of the coefficient vector **c** is obtained by premultiplying both sides of Eq. (2.11) with the pseudo-inverse of operator **GΓΦ**, that is

$$\hat{\mathbf{c}}(\omega;\varpi)=[\mathbf{G}(\omega)\boldsymbol{\Gamma}\boldsymbol{\Phi}(\omega)]^{+}\mathbf{p}(\omega;\varpi)\,. \tag{2.12}$$

This preliminary step, which will be referred to as "backpropagation" in the rest of the dissertation, actually amounts to reconstructing the full source distribution, as it is typically achieved in NAH. As well-known in the literature, this is a severely ill-posed problem which requires very careful regularization in designing the pseudo-inverse $[\mathbf{G}\boldsymbol{\Gamma}\boldsymbol{\Phi}]^{+}$ [[72]-[77], [81]-[86], [93]-[95]]. Although other choices are possible, the Bayesian regularization of Ref. [27] is adopted here due to its good performance and robustness. When used with a Gaussian *prior* for the unknown coefficients, $\mathbf{c}(\omega;\varpi)$, Bayesian regularization returns

$[\mathbf{G}(\omega)\mathbf{\Gamma}\mathbf{\Phi}(\omega)]^{+}=\mathbf{\Phi}(\omega)^{H}\mathbf{\Gamma}\mathbf{G}(\omega)^{H}\left(\mathbf{\Phi}(\omega)^{H}\mathbf{\Gamma}\mathbf{G}(\omega)^{H}\mathbf{G}(\omega)\mathbf{\Gamma}\mathbf{\Phi}(\omega)+\lambda^{2}\right)$ where the regularization parameter λ^2 is set automatically as that value that maximizes the probability of observing the data. The detailed information on the Bayesian regularization is referred to Appendix A.

Now comparing with Eq. (2.11), Eq. (2.12) may be equivalently reformulated as

$$\hat{\mathbf{c}}(\omega;\varpi)=\mathbf{A}(\omega)\boldsymbol{\alpha}(\omega;\varpi)+\mathbf{n}(\omega;\varpi) \tag{2.13}$$

where **n** accounts for estimation noise. The second step is thus to solve Eq. (2.13) for **A** and **α**, which is surely easier than inverting Eq. (2.11) and resembles more the classical BSS problems. This "separation step" is addressed in the next section.

2.3 Decorrelation of sound sources

Statistical decorrelation is often referred to as whitening in BSS, which is the first step of the separation [[42], [96]]. Note that "decorrelation" equals "incoherence" in the present dissertation. With the help of statistical decorrelation, the space of measurements can be divided into two orthogonal subspaces: the signal and the noise subspaces. Meanwhile, the two subspaces can then be represented by two groups of orthogonal vectors, respectively.

2.3.1 Statistical formalization of mutual incoherence

A source is assumed fully coherent with itself and fully incoherent with other sources. This distinctive property is a direct consequence of sources originating from different physical phenomena [57]-[59]. In a probabilistic language, this is expressed as follows. Now, considering all snapshots from measurements, the mutual incoherence between two sources s_i and s_j reads:

$$\frac{\left|\mathbb{E}_{\varpi}\left\{p_i(\mathbf{r}_m,\omega;\varpi)\,p_j^{*}(\mathbf{r}_n,\omega;\varpi)\right\}\right|^2}{\mathbb{E}_{\varpi}\left|p_i(\mathbf{r}_m,\omega;\varpi)\right|^2\,\mathbb{E}_{\varpi}\left|p_j(\mathbf{r}_n,\omega;\varpi)\right|^2}=\delta_{ij},\ \forall(\mathbf{r}_m,\mathbf{r}_n),\ m,n=1,\ldots,M \tag{2.14}$$

where $|\bullet|$ stands for the absolute value. Equation (2.14) mathematically states that

sound sources stemming from different origins are uncorrelated in the space domain, i.e. without any phase relationship [97]. Based on the formulation of mutual incoherence, the source distribution, corresponding to s_i, can be defined as

$$\pi_i(\mathbf{r},\omega) = \mathbb{E}_{\varpi}\left|s_i(\mathbf{r},\omega;\varpi)\right|^2, \tag{2.15}$$

which is a function of the spatial variable **r** in the source domain Γ.

A further working assumption which has its importance in the present work is that the number of microphones in the array, say M, is greater than the number of active sound sources, say N_s. As discussed in the introduction, this is a very realistic assumption in acoustic imaging, where indeed the ratio M/N_s might be much greater than 1 in many applications. For instance, arrays used acoustic imaging typically comprise several tens of microphones to analyze a few units of sources; the example in section 6.3 involve an 84 microphone array (i.e. $M = 84$) to record the sound pressure signals radiated from a Diesel engine, where 2 to 3 sources (combustion noise and injection noise) are of interest ($N_s = 2$).

2.3.2 Decorrelation from eigenvalue decomposition

The aim of BSS is to return an estimate of the mixing matrix, $\hat{\mathbf{A}}(\omega)$, and of the latent variables, $\hat{\boldsymbol{\alpha}}(\omega;\varpi)$, from which the i-th source can be obtained as

$$\hat{\mathbf{s}}_i(\omega;\varpi) = \boldsymbol{\Phi}(\omega)\hat{\mathbf{A}}(\omega)\mathbf{e}_i\mathbf{e}_i^T\hat{\boldsymbol{\alpha}}(\omega;\varpi), \tag{2.16}$$

where $\mathbf{e}_i$ stands for the i-th column of the identity matrix **I** and superscript '^{T}' for the transpose operation. Of primary interest in the following is the estimation of matrix **A**, since it makes possible to estimate the source distribution as

$$\hat{\pi}_i(\mathbf{r}_l,\omega) = \mathbf{e}_l^T\boldsymbol{\Phi}(\omega)\hat{\mathbf{A}}(\omega)\mathbf{e}_i\mathbf{e}_i^T\hat{\mathbf{A}}(\omega)^H\boldsymbol{\Phi}(\omega)^H\mathbf{e}_l \tag{2.17}$$

where the orthogonality property in Eq. (2.7) has been used. This will be achieved thanks to the spatial properties of sound sources, i.e. spatial decorrelation and spatial compactness, as will be introduced in Chapters 3 and 4, respectively. From now on symbol ω standing for angular frequency will be omitted for notational simplification

whenever there is no ambiguity.

The recovery of incoherent sound sources follows a classical approach – see e.g. [58]. Taking into account the statistical decorrelation for the latent variables expressed in Eqs. (2.7) and (2.13), one gets the following result

$$\mathbf{C}_{\hat{\mathbf{c}}\hat{\mathbf{c}}^H} = \mathbb{E}_{\varpi}\left\{\hat{\mathbf{c}}\hat{\mathbf{c}}^H\right\} = \mathbf{A}\mathbf{A}^H + \sigma_n^2\mathbf{I} \tag{2.18}$$

where, for simplicity, the additive noise **n** has been assumed spatially white with variance σ_n^2. In order to recover the mixing matrix **A**, its singular value decomposition,

$$\mathbf{A} = \sum_{i=1}^{N_s} d_i \mathbf{u}_i \mathbf{v}_i^H, \quad d_1 \geq d_2 \geq,\ldots,\geq d_{N_s} > 0 \tag{2.19}$$

is now considered, where $\mathbf{U} = \left[\mathbf{u}_1 \ \cdots \ \mathbf{u}_K\right]$ and $\mathbf{V} = \left[\mathbf{v}_1 \ \cdots \ \mathbf{v}_{N_s}\right]$ are two unitary matrices and $\mathbf{D} = \mathrm{diag}(d_1,\ldots d_{N_s})$ is a nonnegative diagonal matrix. Substituting Eq. (2.19) into Eq. (2.18), it follows that

$$\mathbb{E}_{\varpi}\left\{\hat{\mathbf{c}}\hat{\mathbf{c}}^H\right\} = \sum_{i=1}^{N_s} d_i^2 \mathbf{u}_i \mathbf{u}_i^H + \sigma_n^2 \sum_{i=N_s+1}^{K} \mathbf{u}_i \mathbf{u}_i^H \ , \tag{2.20}$$

which provides a unique solution for matrices **U** and **D** from the eigenvalue decomposition of the covariance matrix of the coefficient vector $\hat{\mathbf{c}}$ (provided that $d_{N_s} > \sigma_n^2$) [98]-[99]. The interpretation of Eq. (2.20) is a decomposition into incoherent components; namely, from Eqs. (2.9) and (2.15), the total source distribution expands as

$$\hat{\pi}(\mathbf{r}_l) = \sum_{i=1}^{N_s} \hat{\pi}_i(\mathbf{r}_l) = \sum_{i=1}^{N_s} d_i^2 \mathbf{e}_l^T \mathbf{\Phi} \mathbf{u}_i \mathbf{u}_i^H \mathbf{\Phi}^H \mathbf{e}_l \ . \tag{2.21}$$

2.4 Virtual sources versus actual sources

Although the above equation (2.21) reflects the conservation of energy (the energy of the source field equals the sum of the energies of incoherent sources), it is not enough to identify the actual source $\hat{\pi}_i(\mathbf{r}_l)$ with the virtual source

$d_i^2 \mathbf{e}_l^T \mathbf{\Phi u}_i \mathbf{u}_i^H \mathbf{\Phi}^H \mathbf{e}_l = \mathbf{e}_l^T \mathbf{\Phi U D e}_i \mathbf{e}_i^T \mathbf{D U}^H \mathbf{\Phi}^H \mathbf{e}_l$. Indeed, a closer comparison with Eq. (2.17), $\hat{\pi}_i(\mathbf{r}_l) = \mathbf{e}_l^T \mathbf{\Phi U D V}^H \mathbf{e}_i \mathbf{e}_i^T \mathbf{V D U}^H \mathbf{\Phi}^H \mathbf{e}_l$ shows that the latter has no reason to be identified with the former since matrix **V** is still missing. This proves that the property of mutual incoherence is not a sufficient condition for separation.

The virtual sources can be considered as orthogonal coordinate axes which span the whole signal space. Obviously, there should be a specific angle which corresponds to the physical system of interest. The specific angle involves certain special information which emphasizes the uniqueness of solutions. Similarly, although they have the same signal space as the virtual sources, the actual sources have distinct properties in terms of spatial distribution which imply their uniqueness. In other words, the two source spaces have equal energy, but have differences in spatial information. Indeed, there exists an infinite number of *incoherent* – but virtual – sources assigned to arbitrary *unitary* matrices **V**. Therefore, the solution to the BSS problem also requires the search for the actual matrix **V** and then the recovery of the actual sources with specified spatial information, which is undertaken from two points of view in Chapters 3 and 4, respectively.

2.5 Blind separation of sources with SOBI

In order to search for the missing matrix **V**, one classic algorithm –SOBI, which only simultaneously combines statistical information at specific frequencies, is concisely introduced here for the blind separation of sound sources [48]. The performance of SOBI will be compared with the proposed methods respectively from Chapters 3 and 4 in Part I. Based on the results from Subsection 2.3 and a group of preselected frequency points with center frequency ω_0, SOBI is executed by the following implementations:

1) Estimate the correlation matrix $\mathbf{C}_{\hat{\mathbf{c}}\hat{\mathbf{c}}^H}(0)$ of the sample coefficient vector $\hat{\mathbf{c}}$ over ϖ snapshots at the center frequency ω_0 as shown in Eq. (2.18). The first N_s largest eigenvalues of $\mathbf{C}_{\hat{\mathbf{c}}\hat{\mathbf{c}}^H}(0)$ and the corresponding eigenvectors are selected as

shown in Eq. (2.19) with the help of EVD.

2) Estimate the noise variance – σ_n^2, as the average of the left (K-N_s) smallest eigenvalues of $\mathbf{C}_{\hat{\mathbf{c}}\hat{\mathbf{c}}^H}(0)$, i.e.

$$\hat{\sigma}_n^2 = \frac{1}{K - N_s} \sum_{i=N_s+1}^{K} d_i^2 , \tag{2.22}$$

when the noise is assumed to be spatially white and to converge to Gaussian white distribution in the frequency domain in terms of CLT [70]. The whitened coefficient vector, $\mathbf{z}(\omega_0) = [z_1(\omega_0), \ldots, z_{N_s}(\omega_0)]^T$, is then determined via

$$z_i(\omega_0) = (d_i^2 - \hat{\sigma}_n^2)^{-(1/2)} \mathbf{u}_i^H \mathbf{c}(\omega_0), \tag{2.23}$$

which is equivalent to establishing an estimate of the whitening matrix, $\hat{\mathbf{W}}$, by

$$\hat{\mathbf{W}} = [(d_1^2 - \hat{\sigma}_n^2)^{-(1/2)} \mathbf{u}_1, ..., (d_{N_s}^2 - \hat{\sigma}_n^2)^{-(1/2)} \mathbf{u}_{N_s}]^H . \tag{2.24}$$

3) Reconstruct a series of covariance matrices $\mathbf{C}_{\hat{\mathbf{c}}\hat{\mathbf{c}}^H}(\nu_j)$ according to a predefined set of frequency lags $\nu \in \{\nu_j \mid j = 0,1, ..., J\}$ with

$$\mathbf{C}_{\hat{\mathbf{c}}\hat{\mathbf{c}}^H}(\nu_j) = \mathbb{E}_{\varpi}\left(\hat{\mathbf{c}}(\omega_0 + \nu_j)\hat{\mathbf{c}}^H(\omega_0 + \nu_j)\right) \tag{2.25}$$

and with $\mathbf{C}_{\hat{\mathbf{c}}\hat{\mathbf{c}}^H}(0)$ when $j = 0$.

4) Search for a unitary matrix $\hat{\mathbf{V}}$ as joint diagonalizer of the set $\{\mathbf{C}_{\hat{\mathbf{c}}\hat{\mathbf{c}}^H}(\nu_j) \mid j = 0,1, ..., J\}$, i.e.

$$\hat{\mathbf{V}} = \underset{\mathbf{V}}{\text{Argmin}}\, \text{off}\left(\sum_{j=0}^{J} \mathbf{W}\mathbf{C}_{\hat{\mathbf{c}}\hat{\mathbf{c}}^H}(\nu_j)\mathbf{W}^H \right) \tag{2.26}$$

where 'off(•)' denotes the sum of squares of all non-diagonal elements of the cost function $\sum_{j=0}^{J} \mathbf{W}\mathbf{C}_{\hat{\mathbf{c}}\hat{\mathbf{c}}^H}(\nu_j)\mathbf{W}^H$. There are many optimization algorithms for the minimization of the cost function in Eq. (2.26). Here, we recommend the algorithm of Joint Approximate Diagonalization of the matrices proposed by J. F. Cardoso [100].

5) Estimate latent sources as

$$\hat{\boldsymbol{\alpha}}(\omega_0) = \hat{\mathbf{V}}\hat{\mathbf{W}}\mathbf{c}(\omega_0), \tag{2.27}$$

and the mixing matrix $\mathbf{A}$ as

$$\hat{\mathbf{A}} = \hat{\mathbf{W}}^{+}\hat{\mathbf{V}}^{H}. \tag{2.28}$$

With the help of the spatial bases $\boldsymbol{\Phi}$ and the estimator $\hat{\mathbf{A}}$, each single sound source could be separated out at ω_0 via Eq. (2.17).

Chapter 3 Blind separation of disjoint sources from spatial decorrelation

In some cases, different physical origins are spatially "orthogonal" to each other, i.e. they have no overlap in space. In the present chapter, a novel way of combing the spatial orthogonality to blindly separate sound sources, is investigated in detail.

3.1 When are virtual sources coinciding with actual sources?

3.1.1 The general case

In Section 2.4, the difference between the virtual sources and the actual sources was discussed in detail. It was shown that decorrelation is generally not a sufficient condition for separating the sources, that is the virtual sources returned by the eigenvalue decomposition of the correlation matrix $\mathbf{C}_{\hat{\mathbf{c}}\hat{\mathbf{c}}^H}$ do not coincide with the actual sources. However, it by no means implies that the virtual sources equal the actual ones in general. To see this, let us compare again the power produced by the *i*-th virtual source as given by Eq. (2.21) (which is reproduced here for convenience),

$$\hat{\pi}_i(\mathbf{r}_l) = \mathbf{e}_l^T \mathbf{\Phi U D e}_i \mathbf{e}_i^T \mathbf{D U}^H \mathbf{\Phi}^H \mathbf{e}_l = d_i^2 \mathbf{e}_l^T \mathbf{\Phi u}_i \mathbf{u}_i^H \mathbf{\Phi}^H \mathbf{e}_l , \tag{3.1}$$

with the actual power as given by Eq. (2.17),

$$\pi_i(\mathbf{r}_l) = \mathbf{e}_l^T \mathbf{\Phi U D V}^H \mathbf{e}_i \mathbf{e}_i^T \mathbf{V D U}^H \mathbf{\Phi}^H \mathbf{e}_l \tag{3.2}$$

where matrix $\mathbf{A}=\mathbf{UDV}^H$ stands for the singular value decomposition of matrix $\mathbf{A}$ (see Eq. (2.19)) with $\mathbf{D}$ an $N_s \times N_s$ diagonal matrix and $\mathbf{U}$ $(M \times N_s)$ and $\mathbf{V}$ $(N_s \times M)$ two unitary matrices such that $\mathbf{UU}^H=\mathbf{VV}^H=\mathbf{I}$. Upon identification, it readily comes that

$$\mathbf{s}_i = \underbrace{\mathbf{\Phi U D}}_{\left[\hat{\mathbf{s}}_1 \ \cdots \ \hat{\mathbf{s}}_{N_s}\right]} \underbrace{\mathbf{V}^H \mathbf{e}_i}_{\begin{bmatrix} v_{1i}^* \\ \vdots \\ v_{N_s,i}^* \end{bmatrix}} = \sum_{k=1}^{N_s} \hat{\mathbf{s}}_k v_{ki}^* , \tag{3.3}$$

where v_{ki}^{*} stands for element (k,i) in matrix $\mathbf{V}^{H}$. This proves that, in general, an actual source is a linear combination of the virtual sources. The coefficients in the linear combination are constrained to conserve energy (i.e. they are the elements of a unitary matrix), but otherwise arbitrary. This leaves infinity of possibilities. The question now arises whether there exists a peculiar physical situation (endowed with an additional constraint) where the virtual sources do equal the actual sources.

3.1.2 An important particular case: spatially uncorrelated sources

A sufficient condition for the virtual sources to equal the actual sources is given hereafter. From the results of the previous section, it is clear that such a condition should satisfy $\mathbf{V} = \mathbf{I}$ or, more generally, $\mathbf{V} = \mathbf{\Pi}$ where $\mathbf{\Pi}$ is the product of a permutation matrix and a diagonal matrix with unit-magnitude complex entries (i.e. every row and column of the matrix involves only one nonzero element with unit norm). In the latter case, the virtual sources are related to the actual sources up to a permutation $\mathbf{\Pi}$ of their indices and are assigned an arbitrary phase φ which does not change their individual powers:

$$\hat{s}_i(\mathbf{r}) = s_{\Pi(i)}(\mathbf{r})e^{J\varphi}, \quad J^2 = -1. \tag{3.4}$$

It is now proved that the scenario corresponds to the case where sound sources are mutually "spatially uncorrelated" in a given domain and the spatial basis $\mathbf{\Phi}$ chosen accordingly. In mathematical language, this means that the scalar products of any pair of two distinct sources with respect to space is nil, i.e.

$$\left\langle s_i, s_j \right\rangle_{\mathbf{M}} = \int_{\mathbf{r}} \int_{\mathbf{r}'} s_i^{*}(\mathbf{r}, \omega; \varpi) s_j(\mathbf{r}', \omega; \varpi) dM(\mathbf{r}, \mathbf{r}') = 0, \quad \forall i \neq j \tag{3.5}$$

with $dM(\mathbf{r}, \mathbf{r}')$ a given metric introduced for sake of generality. Resorting to the discretization scheme introduced in subsection 2.2, this is equivalently expressed as

$$\left\langle s_i, s_j \right\rangle_{\mathbf{M}} = \mathbf{s}_i^{H} \mathbf{M} \mathbf{s}_j = \alpha_i \alpha_j \mathbf{e}_i^{T} \mathbf{A}^{H} \mathbf{\Phi}^{H} \mathbf{M} \mathbf{\Phi} \mathbf{A} \mathbf{e}_j = 0, \quad \forall i \neq j \tag{3.6}$$

or simply

$$\alpha_i \alpha_j \mathbf{e}_i^{T} \mathbf{\Psi}^{H} \mathbf{M} \mathbf{\Psi} \mathbf{e}_j = 0, \quad \forall i \neq j \tag{3.7}$$

in terms of the mode shapes $\mathbf{\Psi} = \mathbf{\Phi A}$, with $\mathbf{M}$ a matrix composed of elements $\Delta M(\mathbf{r}_k, \mathbf{r}_l)$. Typical choices for $\mathbf{M}$ are $\mathbf{M} = \mathbf{\Gamma}$ (with $\mathbf{\Gamma}$ defined in Eq. (2.1)), $\mathbf{M} = \mathbf{\Sigma}^{-1}\mathbf{\Gamma}$ (where $\mathbf{\Sigma}$ is defined as an aperture function in Ref. [27]) or $\mathbf{M} = \mathbf{G\Gamma}$ as in Eq. (2.5) in which case the spatial orthogonality is not assessed in the source domain but in an arbitrary domain to which sources are radiated thanks to the Green function $\mathbf{G}$ (spatial orthogonality then applies to the radiated sources rather than the sources themselves). Since the above equation must be true whatever the value of the latent variables (α_i, α_j) (*spatial* rather *statistical* decorrelation is of concern), this implies that $\mathbf{A}^H\mathbf{\Phi}^H\mathbf{M\Phi A}$ must be a diagonal matrix, that is

$$\mathbf{A}^H\mathbf{\Phi}^H\mathbf{M\Phi A} = \begin{bmatrix} \lambda_1^2 & & \\ & \ddots & \\ & & \lambda_M^2 \end{bmatrix} = \mathbf{\Lambda}^2 \tag{3.8}$$

where the λ_i^2's are non-negative elements (because matrix $\mathbf{A}^H\mathbf{\Phi}^H\mathbf{M\Phi A}$ is symmetric Hermitian). At the same time, substituting $\mathbf{A}$ for its singular value decomposition given in Eq. (2.19) gives

$$\mathbf{A}^H\mathbf{\Phi}^H\mathbf{M\Phi A} = \mathbf{VDU}^H\mathbf{\Phi}^H\mathbf{M\Phi UDV}^H. \tag{3.9}$$

Equating Eqs. (3.8) and (3.9), one has

$$\mathbf{DU}^H\mathbf{\Phi}^H\mathbf{M\Phi UD} = \mathbf{V}^H\mathbf{\Lambda}^2\mathbf{V}, \tag{3.10}$$

which reveals in passing that the unitary matrix $\mathbf{V}$ and the diagonal matrix $\mathbf{\Lambda}^2$ are the eigen-elements of the Hermitian matrix $\mathbf{DU}^H\mathbf{\Phi}^H\mathbf{M\Phi UD}$ (a property that will be exploited in subsection 3.2.2). It is now seen that Eq. (3.6) is satisfied with $\mathbf{V} = \mathbf{\Pi}$ if either

i) $\mathbf{\Phi}^H\mathbf{M\Phi} = \mathbf{I}$ (C.11)

ii) or $\mathbf{\Phi}^H\mathbf{M\Phi} = \mathbf{UD}^+\mathbf{U}^H = \mathbf{A}^+$ (C.12) where $^+$ stands for the pseudo-inverse operator.

Condition i) is obviously more general than condition ii) and can be easily forced by

constructing basis $\mathbf{\Phi}$ accordingly (it is actually satisfied by construction in the optimal setting described in Ref. [27] with $\mathbf{M} = \mathbf{\Sigma}^{-1}$).

To summarize, the virtual sources equal the actual sources (in the sense of Eq. (3.4)) if

1) the sources are spatially uncorrelated in a given domain, i.e. $\mathbf{\Psi}^H \mathbf{M} \mathbf{\Psi}$ is diagonal as specified by Eq. (3.6)
2) and the spatial basis $\mathbf{\Phi}$ is chosen orthonormal in that domain, i.e. $\mathbf{\Phi}^H \mathbf{M} \mathbf{\Phi} = \mathbf{I}$ as specified by Eq. (C.11).

Condition 1) is necessary and sufficient, while condition 2) is only sufficient since it could be replaced by (C.12). The above proof answers the question first raised by S. M. Price and R. J. Bernhard in their seminal paper [58]. As far as the authors known this simple result has never been formulated *explicitly*. Indeed, the literature provides many examples where efforts have been lost vainly to reach the above objective whereas the required assumption obviously did not hold. A particular situation of interest is when sources of are disjoint (i.e. non-overlapping support sets). Then condition (1) is naturally met with $\mathbf{M}$ diagonal. This is investigated in some depth in the next subsection.

3.2 Separation of spatially disjoint sources

The situation with spatially disjoint sources is a particular case of spatially orthogonal sources. Since the disjointing is considered in the source domain rather than in the measurement region, the situation deserves fair practical interest. Indeed, there are many instances where sound sources are likelihood to originate from different (disjoint) spatial regions, even though their radiated fields rapidly overlap in space at some distance from the source domain. In mathematical terms, disjoint sources are characterized by the diagonality of matrix product $\mathbf{\Psi}^H \mathbf{M} \mathbf{\Psi}$ for any diagonal matrix $\mathbf{M}$. An algorithm is designed in this subsection that exploits this property to achieve source separation independently of the choice of the basis $\mathbf{\Phi}$ (i.e. condition 2) of subsection 3.1.2 is relaxed).

3.2.1 Enforcing spatial decorrelation with joint diagonalization

The starting point of the proposed algorithm is Eq. (3.10) where it is seen that the unknown matrix **V** that relates the virtual sources to the actual ones is actually returned as the modal matrix of the matrix product $\mathbf{DU}^H\mathbf{\Phi}^H\mathbf{M}\mathbf{\Phi}\mathbf{UD}$ wherein all quantities are known. Thus, in theory, matrix **V** can be uniquely recovered as that matrix that jointly diagonalises $\mathbf{DU}^H\mathbf{\Phi}^H\mathbf{M}\mathbf{\Phi}\mathbf{UD}$ for a set of candidate metrics **M**. This novel criterion is referred to as "spatial decorrelation", in the following.

Therefore, the proposed algorithm is the following:

1) Statistical decorrelation: estimate the eigen-elements **U** and **D** from the eigenvalue decomposition in Eq. (2.19),
2) Construct a set of matrices $\left\{\mathbf{DU}^H\mathbf{\Phi}^H\mathbf{M}_k\mathbf{\Phi}\mathbf{UD}\right\}_{k=1}^{K}$ for different candidate metrics $\mathbf{M}_k$,
3) Spatial decorrelation: find the unitary matrix **V** that jointly diagonalizes the set of matrices $\left\{\mathbf{DU}^H\mathbf{\Phi}^H\mathbf{M}_k\mathbf{\Phi}\mathbf{UD}\right\}_{k=1}^{K}$,
4) Recover the individual sources as $\mathbf{s}_i = \mathbf{\Phi}\mathbf{UDV}^H e_i$.

It is noteworthy that the algorithm applies to any choice of the spatial basis $\mathbf{\Phi}$. On the contrary, the metrics $\mathbf{M}_k$ should be carefully chosen. Here, the harmonic function is recommended for the metrics $\mathbf{M}_k$. Although other orthogonal polynomials, such as Legendre ones and Chebyshev ones, could be employed here, they suffer from less smoothness on boundaries than the harmonic function does. The less smoothness on the boundaries is less sensitive to the rotation of the missing matrix **V**, which stops more spatial information from being applied in the joint diagonalization of spatial decorrelation. After all, less spatial information means smaller probability to find the missing matrix **V**.

3.2.2 Joint approximate diagonalization

Let us define

$$C_{ij}^{[k]}(\mathbf{V}) = \mathbf{e}_j^T \mathbf{V}\mathbf{D}\mathbf{U}^H \mathbf{\Phi}^H \mathbf{M}_k \mathbf{\Phi}\mathbf{U}\mathbf{D}\mathbf{V}^H \mathbf{e}_i \,. \tag{3.13}$$

The joint diagonalization of the set of weighted spatial correlation coefficients $\left\{\mathbf{C}^{[k]}(\mathbf{V})\right\}_{k=1}^{K}$ may be naturally achieved by minimizing the sum of the squared magnitudes of their off-diagonal elements, that is

$$\begin{aligned} &\hat{\mathbf{V}} = \underset{\mathbf{V}}{\text{Argmin}} \left\{ \sum_{k=1}^{K} \left| \text{Off}\left\{\mathbf{C}^{[k]}(\mathbf{V})\right\} \right|^2 \right\} \\ &s.c. \quad \mathbf{V}\mathbf{V}^H = \mathbf{I} \end{aligned} \tag{3.14}$$

where operator Off zeroes the diagonal elements of a matrix. Fortunately, the cost function in Eq. (3.14) is common in the array signal processing and enjoys several optimization algorithms. Here, two popular algorithms – Joint Approximate Diagonalization(JAD) [100] based on givens rotations and the conjugate gradient method (CG) introduced in Section 4.3 are used.

3.3 How does backpropagation affect the spatial disjoint at low frequency?

According to the theory of backpropagation, the spatial resolution of the reconstructed source distribution at low frequency is much lower than that at high frequency. Lower spatial resolution means lager size of sound sources in the source domain. Therefore, the reconstructed sources at low frequency are apparently larger than their real size and then they overlap to each other in the source domain of interest, although the actual sound sources are spatially disjoint to each other.

As the reconstructed sources overlap to each other at low frequency, the right hand of Eq. (3.8) does not hold any more. The virtual sources from only statistical decorrelation do not equal the actual sources, either. Fortunately, the spatial decorrelation can be still applied to blindly separate the spatially disjoint sources above a specific frequency. Undoubtedly, the lower bound of the working frequency band is determined by the choice of the metrics $\mathbf{M}_k$ and their order K. The effect of backpropagation on the separation of sound sources at low frequency is presented in

Subsection 6.1.2-6.1.5, respectively.

The dependence of spatial decorrelation on backpropagation once again highlights the difficulty of the introduced issue in the present dissertation, and the difference between the incoherent source separation in space and the classic blind source separation.

Chapter 4 Blind separation of compact sources from the principle of least spatial complexity

A specific assumption of the work in the present chapter is that all sound sources of interest are characterized by compact spatial distributions. A source distribution is said compact if it concentrates around a central point, yet without restriction on its shape. Trivial examples are monopoles, dipoles, quadrupoles, but not only. Indeed, the class of compact sources is infinite, despite being restrictive. However, it is surely consistent with many generation mechanisms of sound found in practice (e.g. a specific component on a Diesel engine block) and, more abstractly, with the human perception of a "source".

The postulate is thus that the actual matrix **V** should make the sound sources as spatially compact as possible, among all unitary matrices. In this chapter an important question arises as how to precisely measure spatial compactness. To answer the question, two criteria are proposed here. The first one is that a compact sound source has a low spatial variance about its centroid. The other one is that the compact source has a low entropy, in the sense that most of its energy is concentrated around one or a few points. Based on these two spatial criteria, two implicit cost functions involving matrix **V** are respectively deduced. The optimization strategy of the cost functions is finally introduced.

4.1 Least spatial variance

In statistics, variance is a measure of how far the points in a distribution lie from their mean. Therefore, it is a valid measure of compactness, leading to the "Criterion of Least Spatial Variance" (CLSV) which forces the separation of sound sources to

maximally concentrate around “hot spots”.

The spatial variance Σ_i of the i-th source s_i is defined as

$$\Sigma_i = \int_\Gamma P_i(\mathbf{r}) \left|\mathbf{r} - \mathbf{r}_{0i}\right|^2 d\Gamma(\mathbf{r}), \tag{3.15}$$

where

$$P_i(\mathbf{r}) = \frac{\pi_i(\mathbf{r})}{\pi_i} \tag{3.16}$$

stands for the normalized “intensity” of the source with $\pi_i(\mathbf{r})$ the source distribution introduced in Eq. (2.15),

$$\pi_i = \int_\Gamma \pi_i(\mathbf{r}) d\Gamma(\mathbf{r}) \tag{3.17}$$

is the source “power”, and

$$\mathbf{r}_{0i} = \int_\Gamma P_i(\mathbf{r}) \mathbf{r} d\Gamma(\mathbf{r}) \tag{3.18}$$

is the spatial centroid of the i-th source distribution. Thus the total spatial variance Σ of the mixture of N_s incoherent sound sources reads

$$\Sigma = \sum_{i=1}^{N_s} \pi_i \Sigma_i . \tag{3.19}$$

Substituting Eq. (2.9) into Eqs. (3.15)-(3.19) and discretizing the related integrals as in Eq. (2.3), one arrives at the following cost function,

$$\Sigma(\mathbf{V}) = \sum_{i=1}^{N_s} \sum_{l=1}^{N} \mathbf{e}_l^T \mathbf{\Phi U D V}^H \mathbf{e}_i \mathbf{e}_i^T \mathbf{V D U}^H \mathbf{\Phi}^H \mathbf{e}_l \, | \mathbf{r}_l - \mathbf{r}_{0i} |^2 \, \Delta\Gamma(\mathbf{r}_l), \tag{3.20}$$

in terms of the unknown unitary matrix $\mathbf{V}$, with

$$\mathbf{r}_{0i} = \frac{\sum_{l=1}^{N} \mathbf{e}_l^T \mathbf{\Phi U D V}^H \mathbf{e}_i \mathbf{e}_i^T \mathbf{V D U}^H \mathbf{\Phi}^H \mathbf{e}_l \mathbf{r}_l \Delta\Gamma(\mathbf{r}_l)}{Tr\{\mathbf{\Gamma \Phi U D V}^H \mathbf{e}_i \mathbf{e}_i^T \mathbf{V D U}^H \mathbf{\Phi}^H\}}, \tag{3.21}$$

where $Tr\{\bullet\}$ stands for the trace operator. CLSV then reads

$$\hat{\mathbf{V}} = \underset{\mathbf{V}}{\text{Argmin}} \, \Sigma(\mathbf{V}), \tag{3.22}$$

which, by construction, tends to separate point sources. CLSV has the benefit of simplicity, however it might not be suited to separate compact but multimodal sound sources. The next criterion is designed to solve this shortcoming.

4.2 Least spatial entropy

4.2.1 Introduction to spatial entropy

Entropy, as one of the most popular concepts from physics, originating from the beginning of the 19th century, has widespread applicability: thermodynamics [101], information theory [102], statistical mechanics [103], geographics [104], and economics [105], to name just a few here. Among all available definitions of entropy, Shannon entropy is preferred here as it mainly works for measuring information, choice and uncertainty in data, which is very similar to our scenario but in space [102]. The well-known formula of Shannon entropy is

$$H = -K\sum_{i=1}^{N} p_i \log p_i \,, \tag{3.23}$$

with

$$\sum_{i=1}^{N} p_i = 1 \,, \tag{3.24}$$

where K is a arbitrary positive constant, p_i represents the probability of event i, and N denotes the total number of possible events. Similarly, Shannon also defines the entropy for a continuous distribution with the Probability Density Function (PDF) $p(x)$ by:

$$H = -\int_{-\infty}^{\infty} p(x) \log p(x) dx \,, \tag{3.25}$$

with

$$\int_{-\infty}^{\infty} p(x) = 1 \,. \tag{3.26}$$

Inspired by the Shannon entropy of information theory, an original statistic concept, named ‘spatial entropy’, is proposed here to measure the spatial compactness of sound sources. The spatial entropy corresponding to the actual sources is the minimum among the values of all candidate sources. That is to say, the spatial redundancy of the candidate sources is more than that of actual sources. Thus the uniqueness of the solution to the actual sources is guaranteed by the spatial entropy.

Interestingly, the spatial entropy for sound sources has the same name as that proposed by M. Batty and then applied in geographical analysis in the 70's [106]-[107]. In those two papers, the spatial entropy is utilized to define cutoff points for boundary definitions of partitioned spatial systems. Although the two concepts have the same name and similar mathematical formula, there are two significant differences: 1) in geographical analysis, the spatial entropy is only applied to the real number, while in acoustics, it has to deal with the complex data; 2) the spatial entropy suffers from no limitation in geographic analysis, while here, it needs to take into account the overlap between adjacent sound sources. From now on, the spatial entropy is limited to the one for measuring the spatial compactness of sound sources without supernumerary introduction. The concrete definition will be presented in the following subsection.

4.2.2 The criterion of least spatial entropy

In an analogous manner of Shannon entropy [102] and the spatial entropy in geographical analysis [106]-[107], the spatial entropy H_i of the i-th source is defined as

$$H_i = -\int_\Gamma P_i(\mathbf{r}) \ln P_i(\mathbf{r}) d\Gamma(\mathbf{r}), \tag{3.27}$$

and the total spatial entropy H of the mixture of N_s incoherent sources as

$$H = \sum_{i=1}^{N_s} \pi_i H_i = -\sum_{i=1}^{N_s} \int_\Gamma \pi_i(\mathbf{r}) \ln P_i(\mathbf{r}) d\Gamma(\mathbf{r}). \tag{3.28}$$

Upon substituting Eq. (2.9) into Eq. (3.28) and discretizing the related integrals as in Eq. (2.3), one arrives at the cost function

$$H(\mathbf{V}) = -\sum_{i=1}^{N_s} \sum_{l=1}^{N} \mathbf{e}_l^T \mathbf{\Phi U D V}^H \mathbf{e}_i \mathbf{e}_i^T \mathbf{V D U}^H \mathbf{\Phi}^H \mathbf{e}_l \Delta\Gamma(\mathbf{r}_l) \ln \left(\frac{\mathbf{e}_l^T \mathbf{\Phi U D V}^H \mathbf{e}_i \mathbf{e}_i^T \mathbf{V D U}^H \mathbf{\Phi}^H \mathbf{e}_l \Delta\Gamma(\mathbf{r}_l)}{Tr\{\mathbf{\Gamma \Phi U D V}^H \mathbf{e}_i \mathbf{e}_i^T \mathbf{V D U}^H \mathbf{\Phi}^H\}} \right) \tag{3.29}$$

in terms of the unknown unitary matrix $\mathbf{V}$. The "Criterion of Least Spatial Entropy"

(CLSE),

$$\hat{\mathbf{V}} = \underset{\mathbf{V}}{\text{Argmin}}\, H(\mathbf{V}), \tag{3.30}$$

forces the energy of the separated sound sources to well concentrate around as few as possible points; this encompasses sound sources such as monopoles, dipoles or quadrupoles for instance, but not only; in theory, it tends to favor distributions that are as least complex as possible.

A flowchart of the proposed BSS method is depicted in Fig. 3.1.

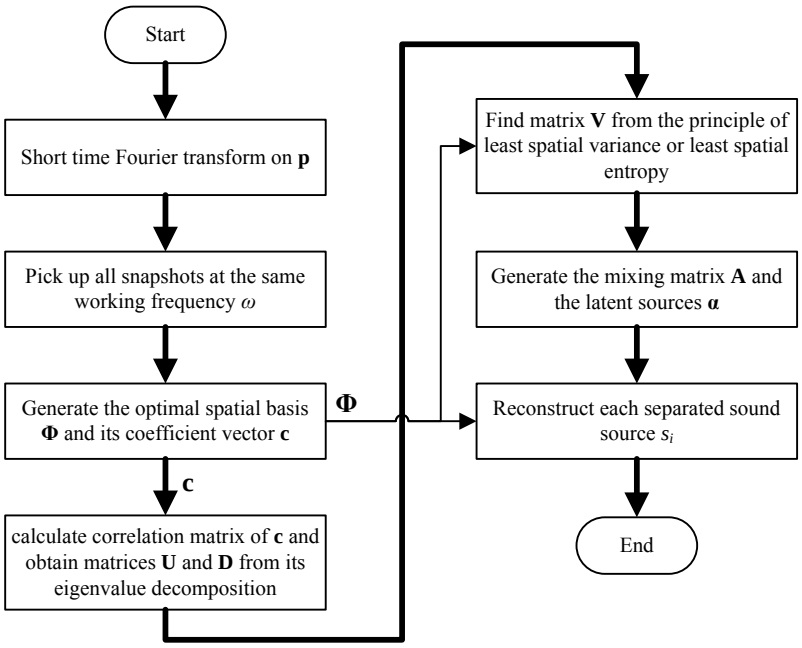

Fig. 4.1 Flowchart of the proposed BSS method.

4.2.3 Ranking eigen elements according to increasing spatial entropy

Traditionally, measurements are divided into two subspaces: the signal subspace and the noise subspace, with eigenvalue decomposition in array signal processing. Most criteria to determine the boundary between the signal subspace and the noise subspace are based on eigenvalues of the covariance matrix of measurements [108].

When there are very small sources (their power is more or less equivalent to that of the noises) embedded in the measurements, the traditional methods might treat such small sources as noise and then throw them into the noise subspace based on their small eigenvalues. That is to say, the very small sources have high probability to be ignored after the operation of statistical decorrelation introduced in subsection 2.3.2.

In order to solve such an intractable problem, new methods need to be proposed from other points of view, e.g. the spatial information, instead of the methods concerning the ranking of eigenvalues. As introduced above, the spatial distribution is a significant inherent attribute for sound sources and it can be considered as a novel criterion to distinguish the source and the noise. As introduced in Section 4.1 and Subsections 4.2.2, the actual sound sources can be treated as a linear combination of all virtual sources which span the whole source space. In the present chapter, the sound sources are all assumed to be compact in space. That means all virtual sources should also be compact to meet the aforementioned assumption. Moreover, the spatial distributions of virtual sources are robust to the size of aperture function. On the other hand, the spatial complexity of the noise will be magnified with the increasing size of aperture function. Inspired by the measures of spatial information, a method based on the spatial entropy (spatial variance could also be applied here) is proposed to separate such small sources.

The fundamental principle is to rank the eigenvalues and their corresponding eigenvectors in terms of the increasing spatial entropies of M virtual sources after the statistical decorrelation. Without loss of generality, the matrix $\mathbf{V}$ for the virtual sources is set to the identity matrix $\mathbf{I}$. Therefore, combining Eqs. (2.16) and (2.20), the m-th virtual source $\bar{\mathbf{s}}_m$ has

$$\bar{\mathbf{s}}_m = d_m \mathbf{\Phi u}_m \bar{\boldsymbol{\alpha}}_m , \tag{3.31}$$

where $\bar{\boldsymbol{\alpha}}_m$ is the m-th latent source corresponding to $\bar{\mathbf{s}}_m$ when $\mathbf{V} = \mathbf{I}$. The spatial entropy of the virtual source $\bar{\mathbf{s}}_m$ could be directly calculated via Eq. (3.27) as

$$H_m = H\left(\bar{\mathbf{s}}_m\right) = -d_m^2 \sum_{l=1}^{N} \mathbf{e}_l^T \boldsymbol{\Phi} \mathbf{u}_m \mathbf{u}_m^H \boldsymbol{\Phi}^H \mathbf{e}_l \Delta\Gamma(\mathbf{r}_l) \ln\left(\frac{\mathbf{e}_l^T \boldsymbol{\Phi} \mathbf{u}_m \mathbf{u}_m^H \boldsymbol{\Phi}^H \mathbf{e}_l \Delta\Gamma(\mathbf{r}_l)}{Tr\{\boldsymbol{\Gamma} \boldsymbol{\Phi} \mathbf{u}_m \mathbf{u}_m^H \boldsymbol{\Phi}^H\}} \right). \quad (3.32)$$

The set of spatial entropies, H_m relating to the M virtual source are reordered in the ascent sequence as $\left\{H_{[m]}\right\}$, where $H_{[m]}$ denotes the m-th smallest spatial entropy. According to the first N_s smallest spatial entropy, the corresponding eigenvalues and eigenvectors are kept respectively as $\bar{\mathbf{D}} = \mathrm{diag}\left(d_{[1]}, \cdots, d_{[N_s]}\right)$ and $\bar{\mathbf{U}} = \left[\mathbf{u}_{[1]} \ \cdots \ \mathbf{u}_{[N_s]}\right]$, and the left ($M$ - N_s) pairs of eigenvalues and eigenvectors are ignored. The selected matrices $\bar{\mathbf{D}}$ and $\bar{\mathbf{U}}$ then join into the search of the missing matrix **V** in the same ways as those presented in Section 3.1 and Subsection 3.2.2, respectively. The validation of the improved method will be illustrated in Part Ⅱ.

From the application point of view, there are two remarks to emphasize. Firstly, we should not only focus on the spatial entropies of virtual sources, but also take care of the corresponding eigenvalues at the same time, when selecting the eigenvalues and eigenvectors of interest for the following optimization. For the actual sound sources, on the one hand, their spatial entropies are much smaller than those from the noise; on the other hand, their energies (or eigenvalues) are not visibly lower than the mean value of the counterpart from the noises. Secondly, to counterpoise calculation efficiency and impressive performance, the size of the aperture function is recommended to be four times of the source domain of interest. After all, too small aperture function will depress the messy spatial distribution of the noises and block the identification between the related virtual sources and the noises.

4.3 Optimization Strategy

Both cost functions (3.22) and (3.30) have to be minimized with respect to the unknown matrix **V**. The fact that **V** is a unitary matrix poses a strong constraint. Though more than ten years has passed, in communication and array signal processing, the problem of optimization under unitary matrix constraint has not to its end. Reference [109] shortly reviews the optimization in manifolds and explains the

disadvantages of the manifold with a Riemannian structure. A possible solution is to transform the constraint optimization problem in the Euclidean space to an unconstraint one in the Stiefel manifold spanned by unitary matrices (e.g. Refs. [110]-[115]).

4.3.1 Optimization in the Stiefel manifold

As mentioned by J. H. Manton [109], Newton algorithm is firstly introduced to optimize a cost function in manifolds. Newton algorithm suffers from the visible limitation – i.e. calculating Hessian of the cost function, although it owns admirable convergence speed – the quadratic convergence. The calculation cost for Hessian might be unaffordable for large matrix [110]. To avoid computing Hessian, SD method is then introduced [[110], [113], [115]]. As we know, SD algorithm takes perpendicular turns at each iteration, which is not a highly efficient way when the cost function is convex but approximately flat near the extremum. Luckily, CG algorithm might significantly improve such a drawback by combining the direction information from the current point and the next one [114]-[115]. Furthermore, CG algorithm converges superlinearly, whereas the convergence of SD algorithm is just linear. Thus CG algorithm is applied herein to optimize the two cost functions – Eqs. (3.22) and (3.30), and search for the optimal unitary matrix $\mathbf{V}$. As presented in Refs [114]-[115], the main steps of CG algorithm in the Stiefel manifold for the two cost functions are summarized as follows:

1) Initialize the index of iteration $k = 0$ and $\mathbf{V}_k = \mathbf{I}$.

2) Calculate the Euclidean gradient $\mathbf{E}_k$ of the cost function – say the total spatial entropy $H(\mathbf{V})$ – with respect to the missing matrix $\mathbf{V}$ is calculated as,

$$\mathbf{E}_k = \frac{\partial H(\mathbf{V}_k)}{\partial \mathbf{V}^H}; \tag{3.33}$$

the corresponding Riemannian gradient $\mathbf{R}_k$ transformed from the Euclidean gradient $\mathbf{E}_k$ in the Stiefel manifold,

$$\mathbf{R}_k = \mathbf{E}_k \mathbf{V}_k^H - \mathbf{V}_k \mathbf{E}_k^H. \tag{3.34}$$

3) Evaluate whether k is the module of n^2 (n a natural number). If it does, the related tangent direction $\mathbf{T}_k$ along the geodesic connecting $\mathbf{V}_k$ and $\mathbf{V}_{k+1}$ becomes $\mathbf{T}_k = \mathbf{R}_k$; otherwise, $\mathbf{T}_k$ is kept.

4) Compute the inner product of the Riemannian gradient $\mathbf{R}_k$ as

$$\left\langle \mathbf{R}_k, \mathbf{R}_k \right\rangle_{\mathbf{I}} = \frac{1}{2} Tr\left\{ \mathbf{R}_k^H \mathbf{R}_k \right\}. \tag{3.35}$$

If the inner product is lower than the predefined threshold value, the whole optimization process will end up immediately.

5) Search for the optimal step size μ_k (which will be introduced in the next subsection).

6) Update the matrix $\mathbf{V}_{k+1}$ as

$$\mathbf{V}_{k+1} = \exp(-\mu_k \mathbf{T}_k)\mathbf{V}_k. \tag{3.36}$$

7) Compute the Euclidean gradient $\mathbf{E}_{k+1}$ and the Riemannian gradient $\mathbf{R}_{k+1}$ at $\mathbf{V}_{k+1}$ in the same way as Eqs. (3.33) and (3.34).

8) Deduce the tangent direction $\mathbf{T}_{k+1}$ at $\mathbf{V}_{k+1}$ as

$$\mathbf{T}_{k+1} = \mathbf{R}_{k+1} + \gamma_k \mathbf{T}_k, \tag{3.37}$$

with the weighting factor

$$\gamma_k = \frac{Tr\{(\mathbf{R}_{k+1} - \mathbf{R}_k)^H \mathbf{R}_{k+1}\}}{Tr\{\mathbf{R}_k^H \mathbf{R}_k\}}. \tag{3.38}$$

9) Evaluate the inner product between the Riemannian gradient $\mathbf{R}_{k+1}$ and the tangent direction $\mathbf{T}_{k+1}$ at $\mathbf{V}_{k+1}$ as

$$\left\langle \mathbf{T}_{k+1}, \mathbf{R}_{k+1} \right\rangle_{\mathbf{I}} = \frac{1}{2} \Re\left\{ Tr\left\{ \mathbf{T}_{k+1}^H \mathbf{R}_{k+1} \right\} \right\}, \tag{3.39}$$

where, $\Re\{\bullet\}$ extracts the real part. Once the inner product becomes negative, the tangent direction $\mathbf{T}_{k+1}$ is forced to be $\mathbf{R}_{k+1}$.

10) Update the index of iteration: $k = k+1$ and go back to step 2).

The key point of successfully applying this algorithm to blind separation of sound sources is how to correctly deduce the Euclidean gradients $\mathbf{E}_k$ of the total spatial variance $\Sigma(\mathbf{V})$ (Eq. (3.22)) and the total spatial entropy $H(\mathbf{V})$ (Eq. (3.30)) with respect to the missing matrix $\mathbf{V}_k$, respectively. The cost functions from the principle of least

spatial complexity are implicit with respect to the mixing matrix **V** as shown in Eqs. (3.22) and (3.30). To be more precise, every single element in the sums of Eqs. (3.22) and (3.30) is not a function of the full matrix **V**, but involves only one of its row.

Unfortunately, the results from the complex-valued matrix gradient of explicit cost functions can not be directly applied here [117]-[122]. Thus, the Euclidean gradients $\mathbf{E}_k$ of the cost functions from the principle of least spatial complexity (see Eqs. (3.22) and (3.30)) needs to be derived carefully, in terms of the original definition of complex-valued matrix gradient of a scalar cost function. The Euclidean gradients $\mathbf{E}_k$ of the total spatial variance $\Sigma(\mathbf{V})$ and the total spatial entropy $H(\mathbf{V})$ are presented in Appendices B and C, respectively.

4.3.2 Optimal step size

Step size is one of the crucial parameters of optimization algorithms, obviously besides the introduced CG algorithm in the Stiefel manifold. The main idea in Refs. [114]-[116] is to search for the optimal step size $\mu_k > 0$ along the geodestic curve

$$\mathbf{\Theta} = \exp\left(-\mu \mathbf{T}_k\right) \mathbf{V}_k , \tag{3.40}$$

which minimizes the cost function, e.g. the total spatial entropy $H(\mathbf{V})$, as

$$\mu_k = \underset{\mu}{\operatorname{Argmin}}\, H(\mathbf{\Theta}) = \underset{\mu}{\operatorname{Argmin}}\, H\left(\exp\left(-\mu \mathbf{T}_k\right) \mathbf{V}_k\right). \tag{3.41}$$

To solve such a problem of minimization, two methods of line search – polynomial approximation and DFT, are proposed in Refs. [114]-[116], respectively. The polynomial approximation approach considers the optimal step size μ_k as the root of the first local minimum of the cost function (i.e. Eq. (3.41)) along a given geodesic. The root of the first local minimum might be too large for the last few iterations and then introduce aggravating oscillation near the minimum of the cost function, which blocks the iteration process finally converging to the optimal value. To avoid such a troublesome problem, a DFT-based method is proposed in Ref [114]-[116]. The goal of the latter approach is to select the best root from the minimum among multiple local minima of the cost function, which corresponds to the minimum of the cost

function in the set. The steps of the two introduced geodesic search approaches – polynomial approximation and DFT, are listed respectively in Tabs. 1 and 2 of Ref. [114] in detail.

Apparently, the optimal step size μ_k from the DFT-based method is much more reasonable than that from the polynomial approximation. As combining more local minima, the DFT-based method becomes more stable at the end of the iterations than the polynomial approximation approach. However, the cost of polynomial approximation algorithm is much lower than that of the DFT-based one. In our case, the calculation of the cost function is very expensive due to combing the optimal spatial basis – a very large matrix of pixel, when working with high spatial resolution. However, the optimal step size from the DFT-based method works much better than that from the polynomial approximation approach. Note that the DFT-based method is recommended, when there are few limitations on the calculation cost.

To conclude and by comparison with Chapter 3, the criterion of spatial decorrelation is much simpler to conceive and to code than the two criteria of spatial compactness (i.e. CLSV and CLSE). The same as the criterion of spatial decorrelation for BSS, the principle of least spatial complexity also needs to determine the number of sources *a priori*. The following chapter will introduce how to deal with the intractable problem in advance of doing BSS.

Chapter 5 Determination of the number of sources

Generally speaking, there are two fundamental problems in array signal processing: 1) determining the number of sources; 2) identifying the positions of sources and analyzing the signal radiated from each source [124]. As a branch of array signal processing, BSS also faces such two problems. To the best of the author's knowledge, in most proposed BSS methods, the number of sources is considered as *a priori* known [[31]-[51], [90]-[92]]. Unfortunately, in most applications, the number of sources is very hard, even impossible, to be known *a priori*. For instance, the number of sound sources radiated from a Diesel engine varies at different frequencies of interest. Therefore, some intelligent methods are required to accurately determine the number of sound sources beforehand, which is not anecdotic: on the one hand, if the source number is underestimated, some sources will be lost in the separation and the remaining ones will be distorted so as to account for the deficit. On the other hand, overestimating the number of sources will introduce spurious sources assigned to noise as well as more calculation efforts in the separation. In this chapter, the currently existing source number determination approaches are firstly surveyed and then four corresponding methods to tackle such a question are investigated.

5.1 Literature survey

In the past 60 years, the problem of determining the number of sources has attracted people from various fields, such as sonar, radar, communication, and geophysics, to name a few hereafter. From the historical point of view, the methods for determining the number of sources experience two primary revolutions.

The first group of methods, named decision theoretic ones [124], becomes popular, after the publishment of [125] and [126]. This kind of methods relies on the statistical

theory of hypothesis testing, mainly concerning the eigenvalues of the spatial correlation matrix. For each hypothesis testing, a predefined appropriate threshold is required. The testing process terminates wherever there exist a likelihood ratio statistic lower than the predefined threshold. However, without enough *a priori* information, it is difficult to accurately set such a subjective but crucial threshold.

Reference [127] introduces a new sphericity test to accurately determine the number of sources present in the acoustic field. The related distribution converges fast to the chi-square, and its performance is comparable with that of the information theoretic criteria (which will be introduced in the next paragraph).

To conquer the subjective limitation from decision theoretic methods and apply more objective information embedded in the eigenvalue spectrum of the spatial correlation matrix, the second group of methods, called information theoretic ones [124], is proposed successively. The sign of its birth is the publication of the famous paper written by M. Wax and T. Kailath in 1985 [128]. The information theoretic method primarily involves two criteria: AIC [129]-[131], and MDL [132]-[134], respectively. Afterwards, many researchers begin to focus on the statistical characteristics of both two information theoretic criteria and their related application scenarios.

L. C. Zhao *et al.* argue that the asymptotic distribution of test statistic of *log* likelihood ratio is not the chi-square; afterwards, they deduce an improved asymptotic distribution for the test statistic [135]. In high resolution array processing, K. M. Wong *et al.* find that AIC overestimates the number of signals at high SNR, but MDL underestimates the parameter at low and medium SNR. To improve the drawbacks of AIC and MDL, they propose a new *log* likelihood function under the framework of information theoretic criteria and modify the corresponding information theoretic criteria, by combining the merits of AIC and BIC [136]. In Ref. [137], W. Xu and M. Kaveh propose a new framework to discuss the exact and approximate asymptotic bounds on the estimations of AIC and MDL, respectively. Their results show that AIC and MDL are both very sensitive to the deviation between the samples and the ideal Gaussian white noise, and that the two criteria overestimate in the presence of colored

Gaussian noise. A. P. Liavas and P. A. Regalia comprehensively analyze the performances of AIC and MDL for the model order selection in different cases – i.e. the large gap between the signal and noise eigenvalues, and the noise eigenvalues dispersion [138]. E. Fishler *et al.* mainly conduct a general asymptotic analysis for MDL in the scenario of estimating the unknown number of Gaussian sources. They claim that the actual distribution of sources does not have much effect on the MDL estimator [139]. In recent years, B. Nadler pays much attention to the detection of signals in the non-parametric setting where no knowledge of the array manifold is known *a priori*. From the viewpoint of random matrix theory, he and his colleague present two new algorithms to determine the number of signals, respectively based on hypothesis testing and information theoretic criteria [140]-[141].

After surveying the existing literatures on determining the number of sources in array signal processing and taking into account the scenario of blind separation of sound sources in both the frequency and space domains introduced in Chapters 2-4, one significant but thorny issue is appears: how to accurately determine the number of sound sources in a simple but efficient way, when the number of sources varies with the working frequency? For instance, in the scenario where the sound sources are radiated from a Diesel engine, the number of significant active sources varies with the frequencies of interest because of the operating conditions and mechanical structures of the engine. To solve the proposed issue effectively, four methods from different mathematical grounds are introduced as follows.

5.2 AIC and MDL

AIC and MDL are the two most well-known methods to detect the number of sources in array signal processing, as introduced in Ref. [128]. AIC and MDL, which are utilized to determine the number of sound sources in the scenarios introduced in Chapters 2-4 at the working frequency ω, are respectively formulated as

$$\mathrm{AIC}\left(\hat{N}_s\right) = -2\left(M-\hat{N}_s\right)N_{\varpi}\log\left(\frac{\prod_{i=\hat{N}_s+1}^{M} d_i^{1/\left(M-\hat{N}_s\right)}}{\frac{1}{M-\hat{N}_s}\sum_{i=\hat{N}_s+1}^{M} d_i}\right) + 2\hat{N}_s\left(2M-\hat{N}_s\right), \qquad (4.1)$$

and

$$\mathrm{MDL}\left(\hat{N}_s\right) = -\left(M-\hat{N}_s\right)N_{\varpi}\log\left(\frac{\prod_{i=\hat{N}_s+1}^{M} d_i^{1/\left(M-\hat{N}_s\right)}}{\frac{1}{M-\hat{N}_s}\sum_{i=\hat{N}_s+1}^{M} d_i}\right) + \frac{1}{2}\hat{N}_s\left(2M-\hat{N}_s\right)\log N_{\varpi}, \qquad (4.2)$$

where $\hat{N}_s$ is the estimated source number and N_{ϖ} denotes the total number of snapshots ϖ.

Generally speaking, MDL works better than AIC. As pointed it out by G. Schwarz, MDL approaches toward the lower-dimensional models due to the penalty term multiplied by $\frac{1}{2}\log N_{\varpi}$ as shown in Eq. (4.2) [132]. In Ref. [142], L. C. Zhao *et al.* prove that the estimation from MDL is consistent, but not for AIC. Furthermore, the number of sources will be definitely determined by AIC, as the number of snapshots, N_{ϖ}, goes to infinity. Therefore, in array signal processing, MDL is preferred over AIC in most cases. However, AIC works much better than MDL in some difficult cases where, for example, no distinct gap exists between the signal and noise eigenvalues [143]. In our case, the number of snapshots, N_{ϖ}, is finite due to the memory capacity of experimental instruments. Thus the estimations from AIC and MDL will be undoubtedly over the actual number of sources due to the limited snapshots, which will be proved by laboratory experiments, by a numerical simulation, and by an industrial example in Chapter6, Part II, respectively. Therefore, other methods for identifying the number of sound sources are necessary.

5.3 The eigenvalue spectrum

The classical approach to determine the number of sources is to inspect the eigenvalue spectrum of the correlation matrix in Eq. (2.20) [144]-[145]. In the high

SNR scenario, there should be a distinct gap between the eigenvalues pertaining the source subspace, $d_1 > d_2 > \ldots > d_{N_s}$, and the (constant) eigenvalues of the noise subspace, $d_{N_s+1} = \ldots d_M = \sigma_n^2$ [[125], [146]]. Unfortunately, this simple criterion is known to perform poorly when additive noise does not follow exactly the working assumptions (spatial whiteness) leading to Eq. (2.20) and the presence of estimation errors [147].

5.4 The cumulative power distribution

The cumulative power of the separated sources may be a useful indicator to determine the number of sources [145]. Under the high SNR scenario, the cumulative power defined as

$$\pi(\hat{N}_s) = \sum_{i=1}^{\hat{N}_s} \pi_i = \sum_{i=1}^{\hat{N}_s} d_i^2 \mathbf{u}_i^H \mathbf{\Phi}^H \mathbf{\Gamma} \mathbf{\Phi} \mathbf{u}_i \tag{4.3}$$

will typically show a "knee" when the estimated number of sources equals the actual number, $\hat{N}_s = N_s$. This might be conveniently inspected by considering the fraction of power, $\pi(\hat{N}_s)/\pi_0$, of the first $\hat{N}_s$ sources to the total power $\pi_0 = \pi(M)$ that can be explained by backpropagation from the M-microphone array.

In practice, one may choose the fraction of power to be no less than 95%, for instance, which will therefore return the minimum number of sources to account for. Differences may therefore exist with the actual source number, but without serious consequence provided the BSS method is robust enough – i.e. provided that the separation of N_1 sources returns more or less the same first N_1 results as the separation of $N_2 > N_1$ sources. This latter assumption will be verified on experimental data in Section 7, Part II.

5.5 The entropic L-curve

Inspired by the principle of the L-curve (as used for determining the optimal

regularization parameter in inverse problems [148]-[150]), a novel approach coined *entropic L-curve*, is proposed hereafter. It is based on the concept of spatial and statistical entropies. Namely, the "solution size" is represented by the total spatial entropy *H* of the separated sources as given by Eq. (3.29), whereas the "residual size" is measured by the Kullback-Leibler divergence [151], *L*, between the full covariance matrix $\mathbf{C}_{\hat{\mathbf{c}}\hat{\mathbf{c}}^H}$ in Eq. (2.18) and its truncated version to $\hat{N}_s$ sources, $\hat{\mathbf{C}}_{\hat{\mathbf{c}}\hat{\mathbf{c}}^H}(\hat{N}_s)$, in Eq. (2.20), i.e.

$$L = Tr\left\{\mathbf{C}_{\hat{\mathbf{c}}\hat{\mathbf{c}}^H}\hat{\mathbf{C}}^{-1}_{\hat{\mathbf{c}}\hat{\mathbf{c}}^H}(\hat{N}_s)\right\} + \ln\frac{\left|\hat{\mathbf{C}}_{\hat{\mathbf{c}}\hat{\mathbf{c}}^H}(\hat{N}_s)\right|}{\left|\mathbf{C}_{\hat{\mathbf{c}}\hat{\mathbf{c}}^H}\right|} - M \ , \tag{4.4}$$

with

$$\begin{aligned}
&\hat{\mathbf{C}}_{\hat{\mathbf{c}}\hat{\mathbf{c}}^H}\left(\hat{N}_s\right) = \sum_{i=1}^{\hat{N}_s}\hat{d}_i^2\mathbf{u}_i\mathbf{u}_i^H + \hat{\sigma}_n^2\sum_{i=\hat{N}_s+1}^{K}\mathbf{u}_i\mathbf{u}_i^H \ , \\
&\hat{d}_i^2 = d_i^2 - \hat{\sigma}_n^2 \ , \ i = 1,\ldots,N_s \ , \\
&\hat{\sigma}_n^2 = \frac{1}{K-\hat{N}_s}\sum_{i=\hat{N}_s+1}^{K}\hat{d}_i^2 ,
\end{aligned} \tag{4.5}$$

where $|\cdot|$ denotes the determinant of a matrix and ln is the natural logarithm. The Kullback-Leibler divergence measures the statistical (relative) entropy of the residual error in fitting the covariance matrix with a given number of sources; therefore it decreases with $\hat{N}_s$ and ultimately vanishes when $\hat{N}_s$=*M* . On the other hand, the total spatial entropy of the separated sources increases with $\hat{N}_s$. Thus the shape of the curve generated by the trajectory of the source number in the plane (*H*, *L*) resembles an inverted "L", where the corner points to the optimum value of $\hat{N}_s$ in the sense of achieving the best compromise between not too large a statistical entropy of the residual error and not too large a spatial entropy of the reconstructed source (since compact sources are to be reconstructed). The idea of balancing between two entropic measures, the first one statistical and the second one spatial, seems to be investigated here for the first time. It is conceptually attractive since it compares two quantities with similar dimensions that account for the *shape* of the reconstructed sources and not only for their energy. Note that in order to keep the computational load reasonable,

it is recommended to compute the spatial entropy H by using the proposed CLSV which is much faster to converge than – even if not as accurate as – CLSE.

As a final remark to this section, it should be emphasized again that determining the number of active sources is a difficult problem which may not be solved by any universal method. Therefore the recommendation is to use the four proposed strategies jointly and to take a decision from their comparison, rather than to adhere to one of them only.

Part II: Experiments

Chapter 6 Experiment validation

The algorithms proposed in Chapters 3-5 are illustrated here on laboratory experiments, numerical simulation, and an industrial example, successively. The performance of all algorithms is discussed in details.

6.1 Laboratory experiments

6.1.1 Experimental apparatus

In order to validate the proposed separation algorithm, a series of experiments were conducted in a semi-anechoic chamber. The experimental apparatus is depicted in Fig. 6.1. A slice wheel array (also named Pizza array) with 60 microphones was placed parallel to four loudspeaker membranes considered as the sound sources, at different distances ranging from the near-field to the far-field [152].

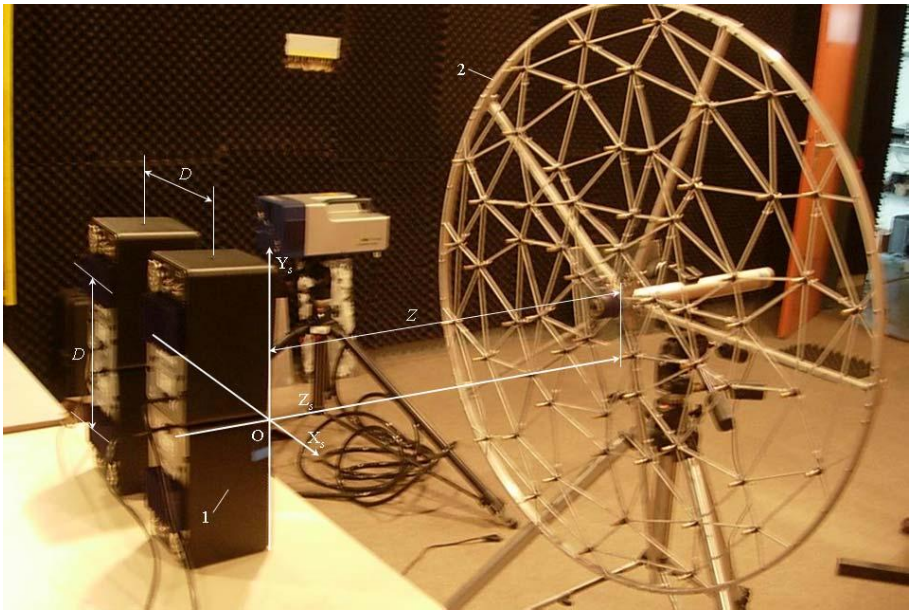

Fig. 6.1 The experimental setup for validating the proposed algorithm. (1: the loudspeakers; 2: the slice wheel microphone array)

The distance between the microphone array and the loudspeaker membranes is

symbolized as Z and the spacing between the centers of two adjacent loudspeaker membranes as D. The combination of values of Z and D is listed in Tab. 1. The number of active loudspeakers – corresponding to the source number N_s – was set to 4 throughout all experiments. All loudspeakers were driven with mutually uncorrelated random white noises via a multi-channel amplifier. The sampling frequency was set to 16.384 kHz and the recording time to 4 min.

In a preliminary step, the loudspeakers were switched on one by one – with the 3 other ones switched off – in order to get the actual source contributions. The latter will serve as point of validation of the blind separation results to be presented hereafter, obtained when all loudspeakers are operating together (the authors insist that such references will not be available in practice, which is purposely the situation that motivates the use of BSS). The corresponding experimental process is illustrated in Fig. 6.2.

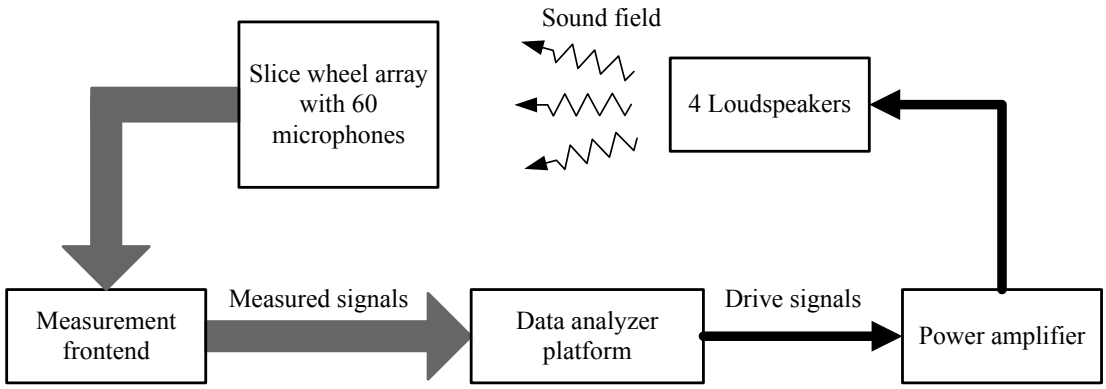

Fig. 6.2 Scheme of the experimental process.

During processing the 60 channel signals, the length of the snapshots was set to 4s; snapshots were further tapered with Hanning windows with 50% overlap. Their number was set to 119.

Z [cm]	10	50	100	200
D [cm]	12	18	24	

Tab. 6.1 Experimental values for distance Z and spacing D.

6.1.2 Separation from single statistical decorrelation

Data were first processed by the simplest separation approach, i.e. forcing Statistical Decorrelation (SD). As discussed in section 4.1, only forcing the statistical decorrelation of the sources by diagonalizing the correlation matrix of coefficient vector **c** can provide satisfactory separation results provided the sound sources have spatially orthogonal distributions.

In the near-field configuration, the distance Z was set to 10 cm, and the spacing D to 12 cm. In the far-field, the two parameters were selected as Z = 100 cm and D = 18 cm.

To present the separation results in the whole working frequency band, generally speaking, power spectrum is the first choice. In the dissertation, the power spectrum is expressed by quadratic velocity spectrum. Unfortunately, the quadratic velocity spectrum only provides energy information for the separation results. It needs another variable to describe the spatial distribution of separated sources in the whole frequency band of interest. Thus, separation results will be displayed by means of the spatial correlation spectrum between the separated and actual sources (i.e. correlation coefficient as a function of frequency), defined as

$$\rho_i(\omega) = \frac{\sum_{l=1}^{N} \left| \mathbb{E}_{\varpi} \left\{ \hat{s}_i(\mathbf{r}_l, \omega; \varpi) s_i^*(\mathbf{r}_l, \omega; \varpi) \right\} \right|}{\sqrt{\sum_{l=1}^{N} \mathbb{E}_{\varpi} \left| \hat{s}_i(\mathbf{r}_l, \omega; \varpi) \right|^2 \sum_{l=1}^{N} \mathbb{E}_{\varpi} \left| s_i^*(\mathbf{r}_l, \omega; \varpi) \right|^2}}, \tag{5.1}$$

where, s_i denotes the i-th actual source, $\hat{s}_i$ is its estimate and * stands for complex conjugate.

1) **Source separation in the near-field**

Preliminary to source separation, the assessment of the source number is carried out as shown in Figs. 6.3 and 6.4, respectively. As shown in Fig. 6.3, information theoretic criteria could not provide a satisfactory number of sound sources as, which

is reminiscent to the discussion of Section 5.2. The minima of AIC and MDL both correspond to $\hat{N}_s = 12$, which largely overestimated the actual number of sources used in the experiement. On the contrary, the eigenvalue spectrum in Fig. 6.4(a), the cumulative power distribution in Fig. 6.4(b) and the entropic L-curve in Fig. 6.4(c) all return a valid estimate of the actual source number: the cumulative power distribution exhibits a corner at $\hat{N}_s = 4$ corresponding to 99.77% of the overall power, indicated in Fig. 6.4(b) by a red arrow. Note that the selected point in Fig. 6.4(c) clearly illustrates the trade-off between the spatial and statistical entropic measures.

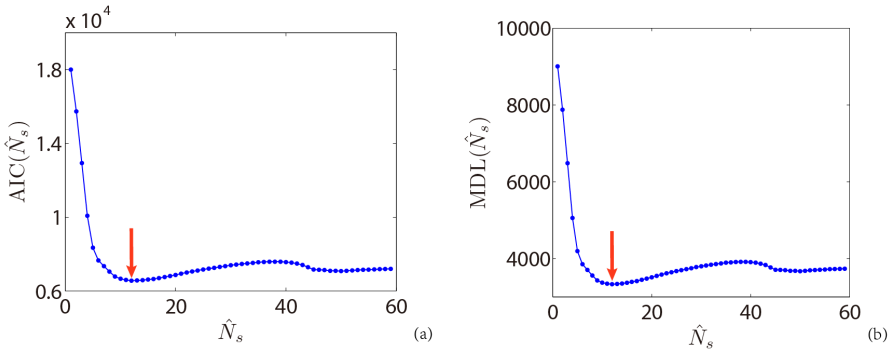

Fig. 6.3 Determination of the source number N_s from inspection of (a) AIC and (b) MDL in the near-field at 833 Hz.

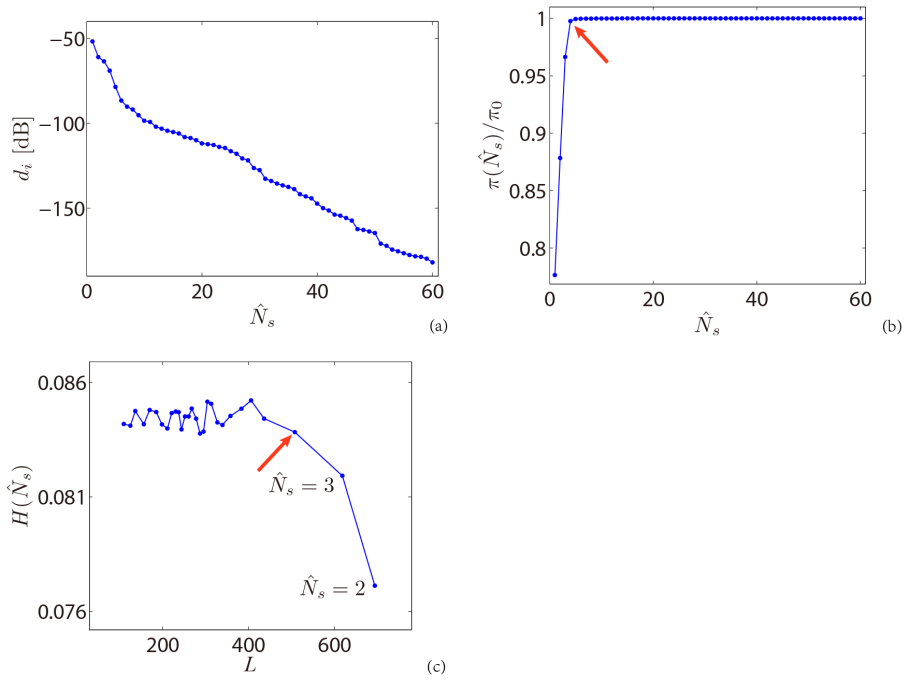

Fig. 6.4 Determination of the source number N_s from inspection of (a) the eigenvalue spectrum, (b) the cumulative power distribution, and (c) the entropic L-curve in the near-field at 833 Hz.

In the near-field, the working frequency is set to 1800 Hz. The separation results from SD are depicted in Fig. 6.5. It appears that all virtual sources account for the sound power and that they could be successfully separated because, at this frequency, their supports are mainly disjoint, a sufficient condition for spatial orthogonality. To evaluate the separation performance, the 4 actual sources, in the near-field at 1800 Hz, are displayed in Fig. 6.6. Ignoring the permutation among the 4 sources (which is a common problem in BSS) and comparing Figs. 6.5-6.6, there are few biases between the virtual sources and the actual ones no matter in amplitude, in location of hottest points, or in spatial distribution.

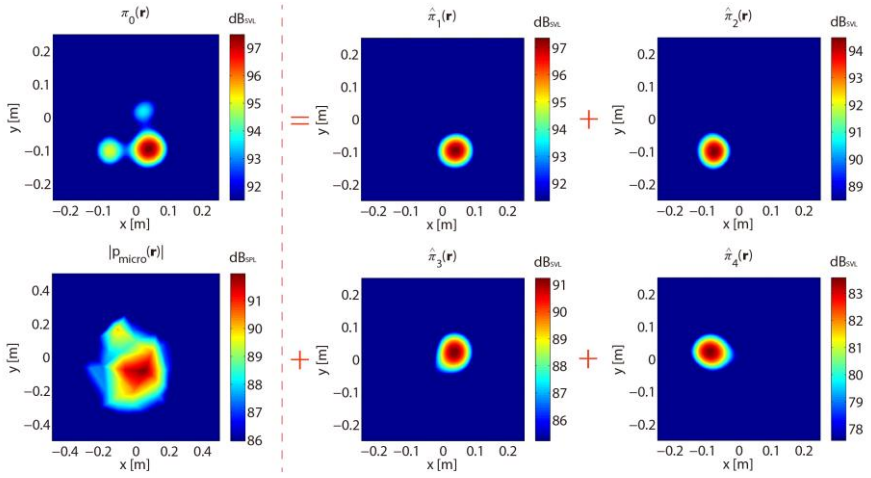

Fig. 6.5 Separated sources, $\hat{\pi}_i(\mathbf{r})$, $i = 1,\ldots,4$, from single statistical decorrelation , in the near-field at 1800 Hz with the aperture function $R = 35$ cm.

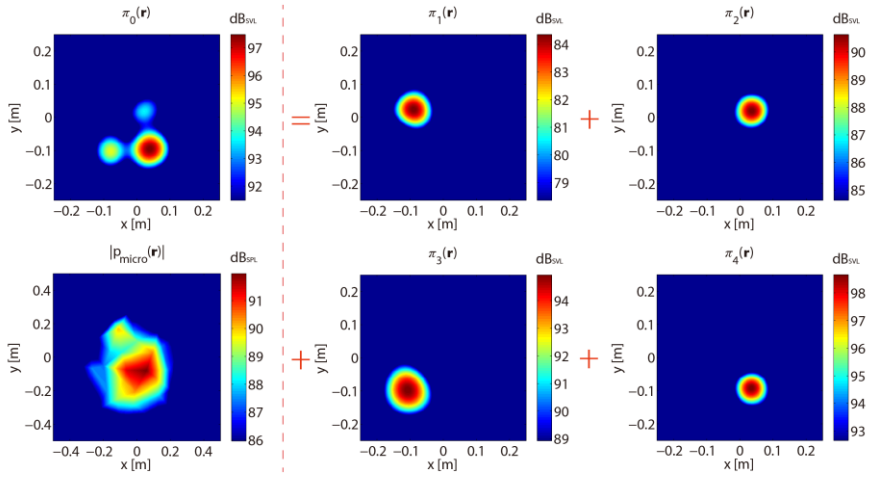

Fig. 6.6 Actual sources, $\pi_i(\mathbf{r})$, $i = 1,\ldots,4$, from individual measurements, in the near-field at 1800 Hz with the aperture function $R = 35$ cm.

Next, the separation results are displayed in terms of quadratic velocity spectra in Fig. 6.7 and spatial correlation spectra in Fig. 6.8 in the whole available frequency band. In order to solve the permutation ambiguity inherent to BSS (arbitrary labeling of the separated sources at each frequency), the quadratic velocity spectra and the spatial correlation spectra were re-ordered at each frequency according to the spatial correlation between the separated sources and the actual ones. The lower bound of the spectra, i.e. the quadratic velocity and the spatial correlation, is determined by the size of microphone array, for instance the diameter of the slice wheel microphone array as shown in Fig. 6.1. On the contrary, the corresponding upper bound depends on the minimum spacing between two adjacent microphones. Neglecting the extremely small sources, for instance, in the frequency band [1300, 1500] Hz that encloses a zero in the quadratic velocity spectrum of the actual source s_1, the energy of the separated sources are well consistent with that of the actual sources except in the very low frequency interval [400, 600] Hz. However, as shown in the spatial correlation spectra (see Fig. 6.8), there exists some discrepancy between the estimated and actual sources even though their energy distributions seemed to match well. The highest spatial correlation – above 0.9 – is mostly at high frequencies (say over 1600 Hz) where the source supports become more disjoint due to finer and finer spatial resolution. The

coefficients of spatial correlation between the separated sources and their corresponding references are 0.99, 0.98, 0.98 and 1 at 1800 Hz (see Fig. 6.8(a)-(d)), respectively.

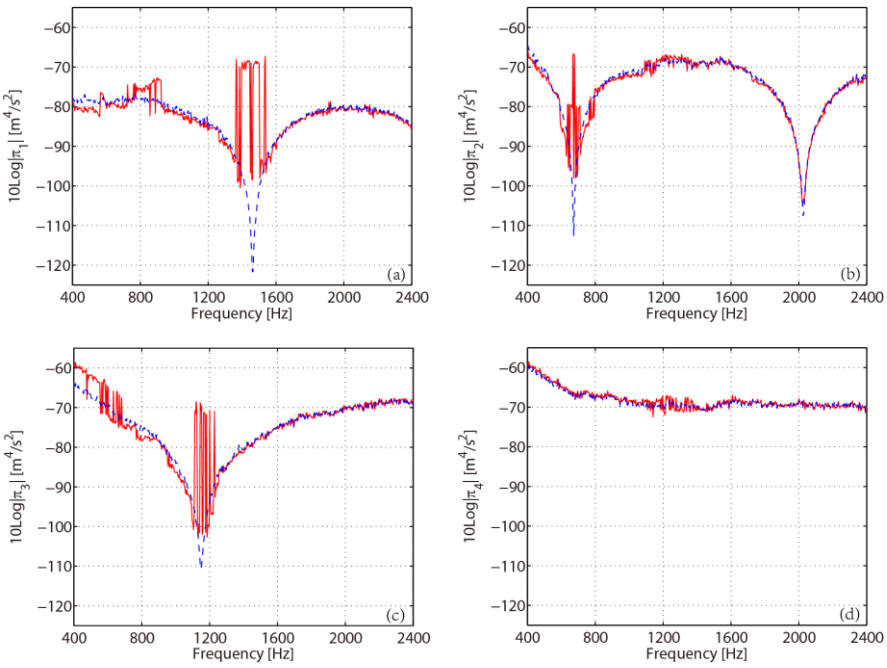

Fig. 6.7 Quadratic velocity spectra of the separated sources, $\hat{\pi}_i(\omega)$, $i = 1,\ldots,4$, from single statistical decorrelation (red solid line) compared to actual sources (blue dashed line), in the near-field.

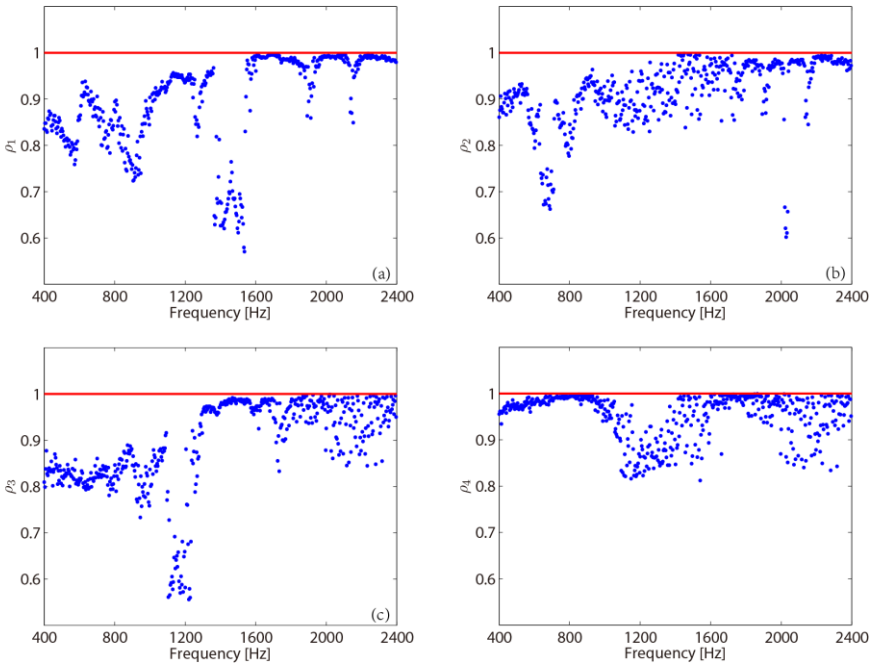

Fig. 6.8 Spatial correlation spectra between the separated and actual sources, $\rho_i(\omega)$, i = 1,…,4, in the near-field.

2) **Source separation in the far-field**

Interestingly, in this case AIC and MDL both exhibit an inflection at the correct number of sources, $\hat{N}_s = 4$, as shown in Fig. 6.9, although this does not correspond to a minimum. The determination of the source number from the other three methods in the far-field is illustrated in Fig. 6.10. Figure 6.10(a) evidences an explicit gap between the eigenvalues of the source and noise subspaces around the actual source number, $N_s = 4$. Meanwhile, this corresponds to a turning point in the power distribution that accounts for 99.97% of the overall power as depicted in Fig. 6.10(b). The corner point in the entropic L-curve is about the same value, as shown in Fig. 6.10(c).

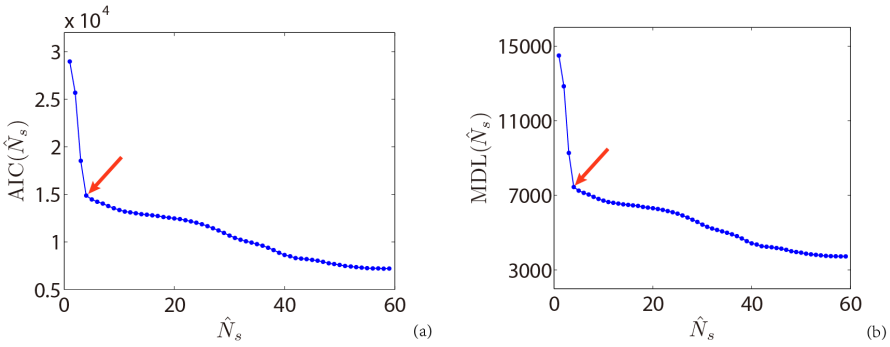

Fig. 6.9 Determination of the source number N_s from inspection of (a) AIC and (b) MDL in the far-field at 2437 Hz.

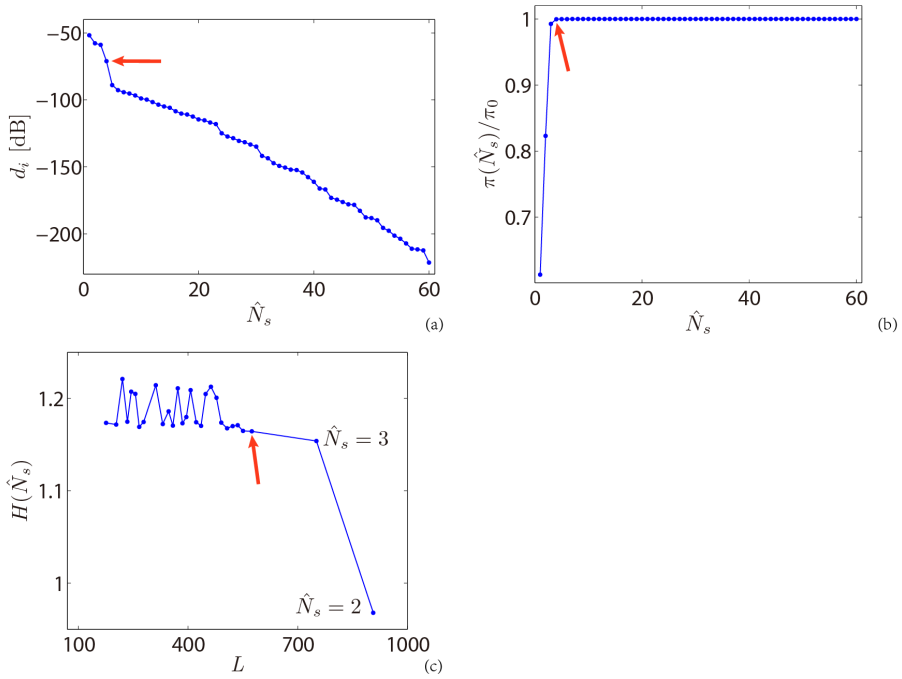

Fig. 6.10 Determination of the source number N_s from inspection of (a) the eigenvalue spectrum, (b) the cumulative power distribution, and (c) the entropic L-curve in the far-field at 2437 Hz.

The separated and actual sources are displayed in Figs. 6.11-6.12, respectively. Separation results are very satisfactory taking into account the amplitudes, the locations of hottest points, and the spatial distributions. This demonstrates that single statistical decorrelation can achieve good spatial separation provided the sources are spatially orthogonal, a condition which is naturally satisfied here when their supports

become disjoint.

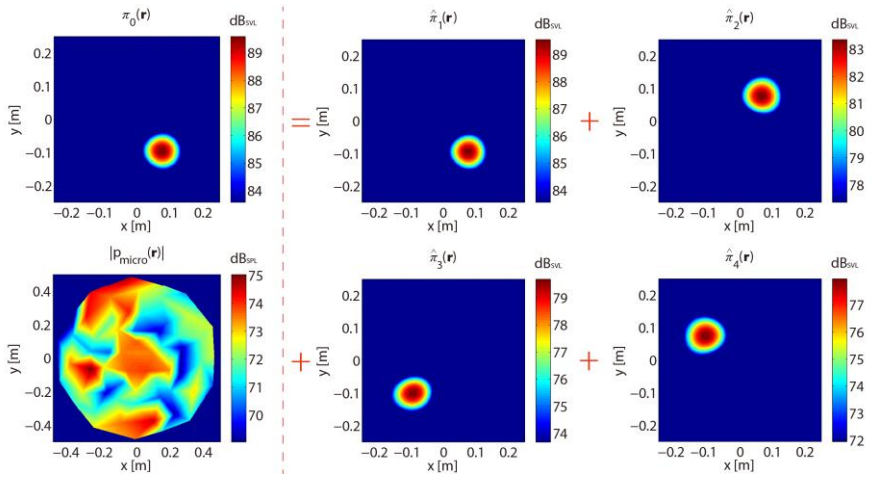

Fig. 6.11 Separated sources, $\hat{\pi}_i(\mathbf{r})$, $i = 1,\ldots,4$, from single statistical decorrelation , in the far-field at 3600 Hz with the aperture function $R = 35$ cm.

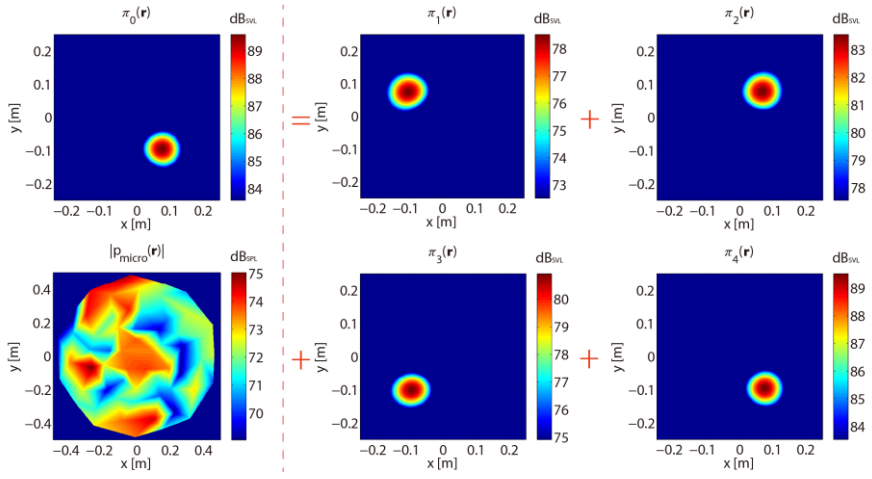

Fig. 6.12 Actual sources, $\pi_i(\mathbf{r})$, $i = 1,\ldots,4$, from individual measurements, in the far-field at 3600 Hz with the aperture function $R = 35$ cm.

Figures 6.13-6.14 display the quadratic velocity spectra and the spatial correlation spectra in the whole available frequency band, respectively. Similar to that in the near-field, the boundaries of the whole working frequency band in the far-field are also determined by characteristic parameters of the microphone array and the distance Z. The quadratic velocity spectra of the separated sources match remarkably well

those of the actual sources, even at the zeros of the first source (see Fig. 6.13(a)). The reason of such an impressive coincidence in the quadratic velocity is due to the larger spacing – D = 18 cm used in the far-field. However, the spatial correlation spectra present distinct results from the quadratic velocity spectra. It is clear that above 3200 Hz, the separated sources have high spatial correlation with the actual ones. In the whole work frequency band, the spatial correlation of the separated sources with success is basically over 0.98. That means the separated sources from single statistical decorrelation are more reliable with wider spacing between actual sources. Corresponding to Figs. 6.11-6.12, the coefficients of the spatial correlation are 0.98, 1, 0.96 and 1 (see Fig. 6.14(a)-(d)), respectively.

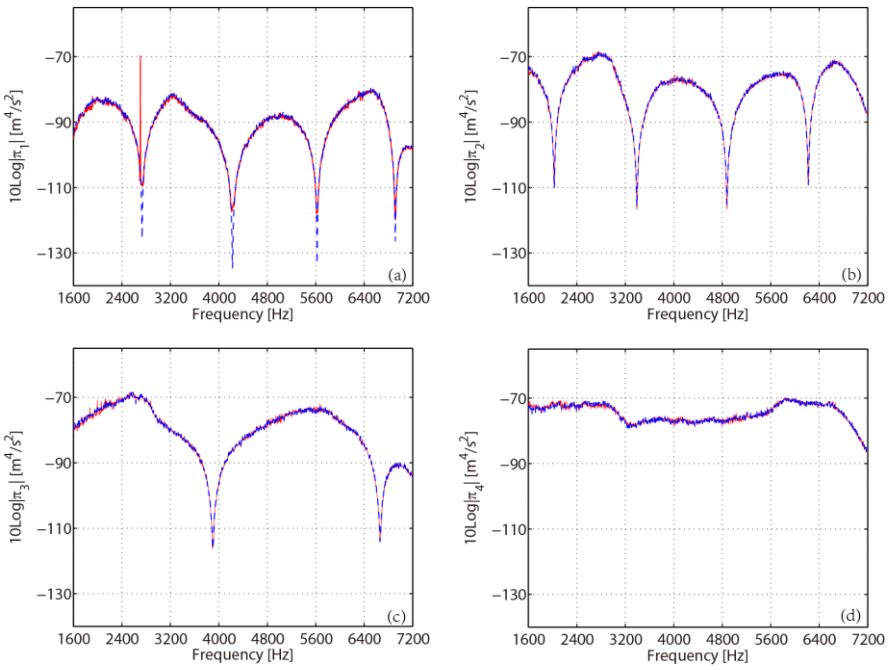

Fig. 6.13 Quadratic velocity spectra of the separated sources, $\hat{\pi}_i(\omega)$, i = 1,…,4, from single statistical decorrelation (red solid line) compared to actual sources (blue dashed line), in the far-field.

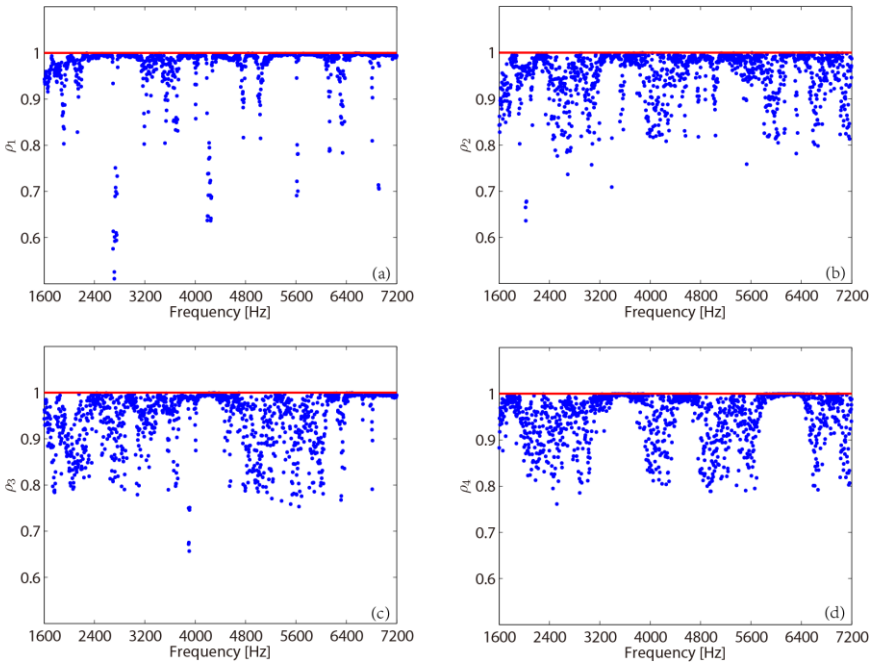

Fig. 6.14 Spatial correlation spectra between the separated and actual sources, $\rho_i(\omega)$, i = 1,…,4, in the far-field.

6.1.3 Separation from joint statistical and spatial decorrelation

When sound sources are disjoint in space, mainly at low and medium frequencies, they might be separated by enforcing Statistical and Spatial Decorrelation (2SD) together with joint approximate diagonalization (see Section 3.2). The experimental parameters of this subsection are the same as in Subsection 6.1.2.

1) **Source separation in the near-field**

To allow comparisons with the separation results of Subsection 6.1.4, the working frequency is chosen as 833 Hz. The separation results from single statistical decorrelation are depicted in Fig. 6.15 with an aperture function R = 35cm. This shows how the particle velocity field (normal component to the source plane), $\pi_0(\mathbf{r},\omega)$, reconstructed from backpropagation of the measured pressures can be decomposed into 4 incoherent sources. It appears that one virtual source, $\hat{\pi}_1(\mathbf{r},\omega)$,

accounts for most of the sound power and that the other sources could hardly be separated by enforcing statistical decorrelation only. The reason is that at such a low frequency the spatial resolution is no longer sufficient to guarantee disjoint supports and thus spatial orthogonality. As explained in Subsection 2.4, this is no surprise since an infinite number of virtual sources could explain equally well the source field as long as the exact value of the unitary matrix **V** is not known.

To improve the separation results, spatial decorrelation is then enforced on the virtual sources. Two optimization strategies: JAD of the matrices [100] and CG [114]-[115] in the Stiefel manifold, are applied to search for the missing matrix **V**. The separated results from the two strategies are illustrated in Figs. 6.16-6.17, respectively. Both optimization strategies return correct and compact sound sources as shown on the right side of the red dash line. The separation results from the two strategies look the same no matter the amplitude, the location of hottest points or the spatial distribution, ignoring the permutation among the four sources. The actual sources at 833 Hz with R = 35 cm are displayed in Fig. 6.18. There are few errors between the separated and actual sources in terms of energy and spatial distribution, although the shape of the second sound source is distorted a little. The example shows that enforcing statistical and spatial decorrelation jointly is an effective method to blindly separate incoherent sound sources in the near-field.

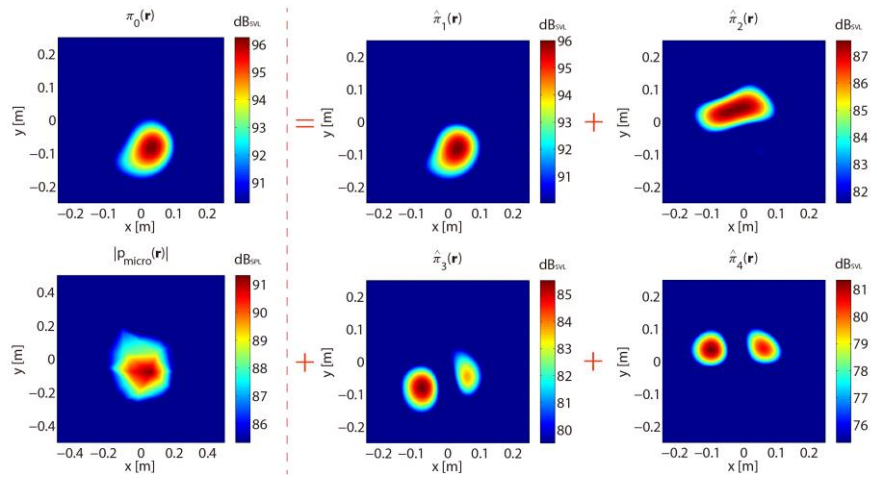

Fig. 6.15 Separated sources, $\hat{\pi}_i(\mathbf{r})$, i = 1,…,4, from single statistical decorrelation, in the

near-field at 833 Hz.

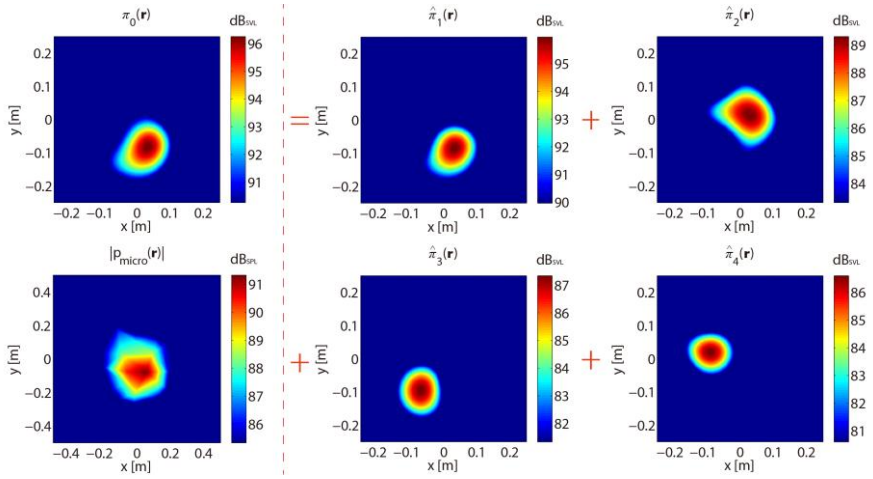

Fig. 6.16 Separated sources, $\hat{\pi}_i(\mathbf{r})$, $i = 1,\ldots,4$, from statistical and spatial decorrelation with joint approximate diagonalization, in the near-field at 833 Hz.

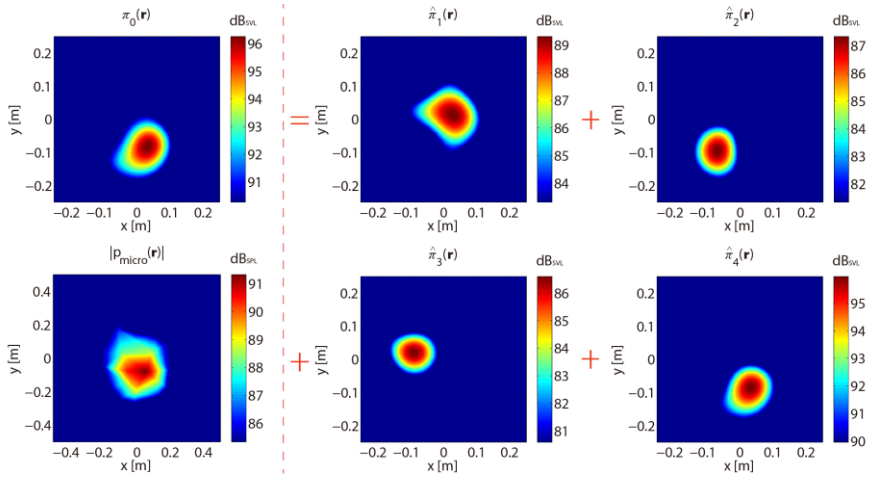

Fig. 6.17 Separated sources, $\hat{\pi}_i(\mathbf{r})$, $i = 1,\ldots,4$, from statistical and spatial decorrelation with conjugate gradient method, in the near-field at 833 Hz.

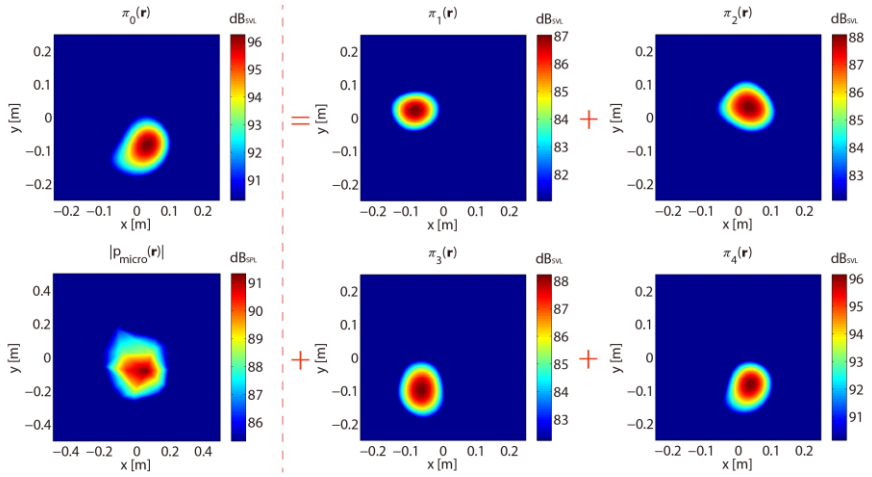

Fig. 6.18 Actual sources, $\pi_i(\mathbf{r})$, $i = 1,\ldots,4$, from individual measurements, in the near-field at 833 Hz.

The quadratic velocity and spatial correlation spectra in the whole available frequency band are shown in Figs. 6.19-6.22, respectively. Comparing Figs. 6.19-6.20 with Fig. 6.7, the quadratic velocity spectra from 2SD degenerate a little relative to those from SD, as there are small differences between the rotation matrices from SD and 2SD. Concerning the two strategies, the quadratic velocity spectra from CG seem in some sort better than those from JAD, mainly on the second zero of the second actual source (see Fig. 6.19(b) and Fig. 6.20(b)).

The spatial correlation spectra in Figs. 6.21-6.22 show improved separation sources in the whole working frequency band as compared to single statistical decorrelation, and ignoring the very small levels near the zeros of the quadratic velocity spectra. Remarkably, the spatial correlations on the third and fourth actual sources are more than 0.98 at most frequencies in the working band (see Fig. 6.21(c)-(d)).

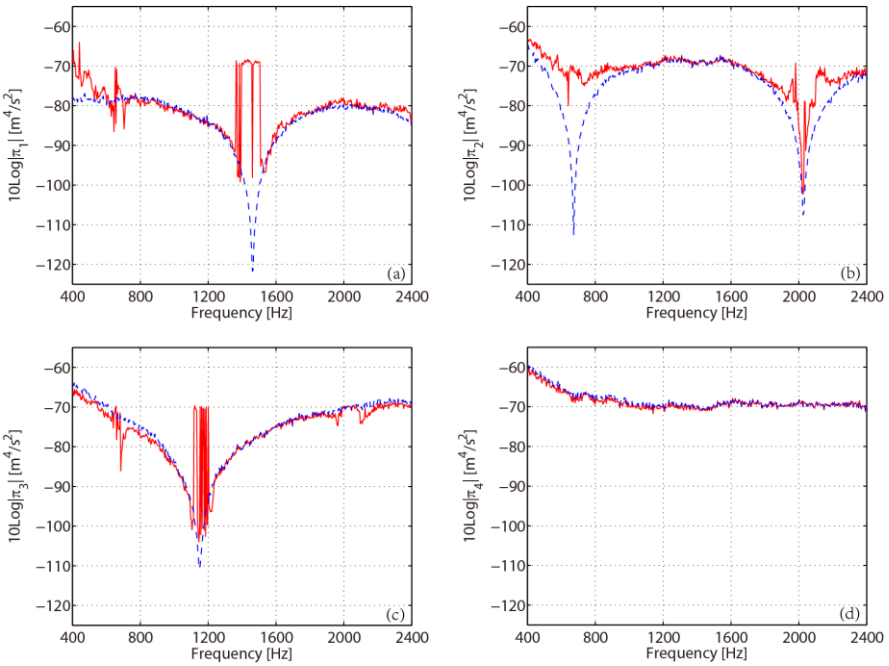

Fig. 6.19 Quadratic velocity spectra of the separated sources, $\hat{\pi}_i(\omega)$, $i = 1,\ldots,4$, from statistical and spatial decorrelation with joint approximate digonalization (red solid line) compared to actual sources (blue dashed line), in the near-field.

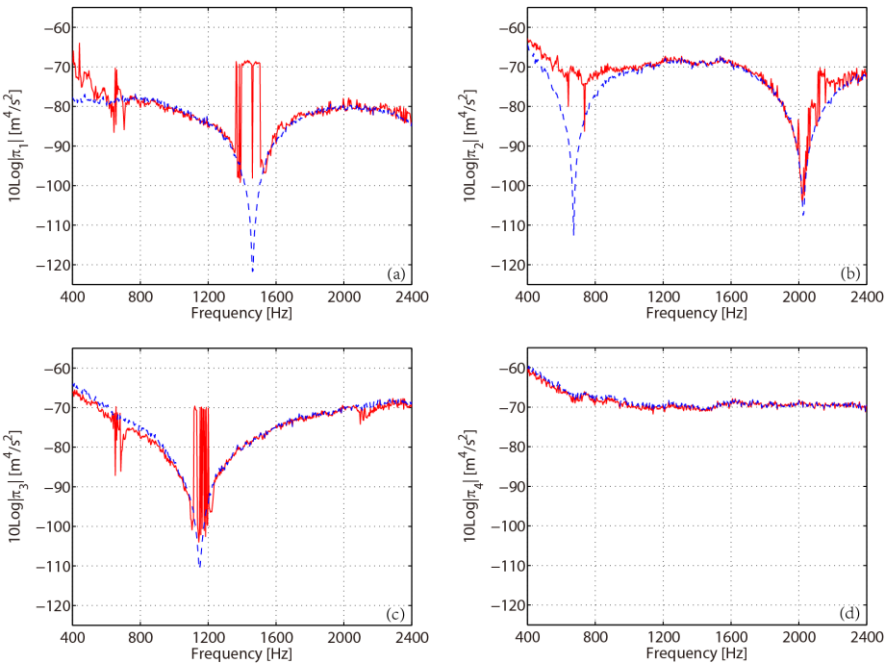

Fig. 6.20 Quadratic velocity spectra of the separated sources, $\hat{\pi}_i(\omega)$, $i = 1,\ldots,4$, from statistical and spatial decorrelation with conjugate gradient method (red solid line) compared

to actual sources (blue dashed line), in the near-field.

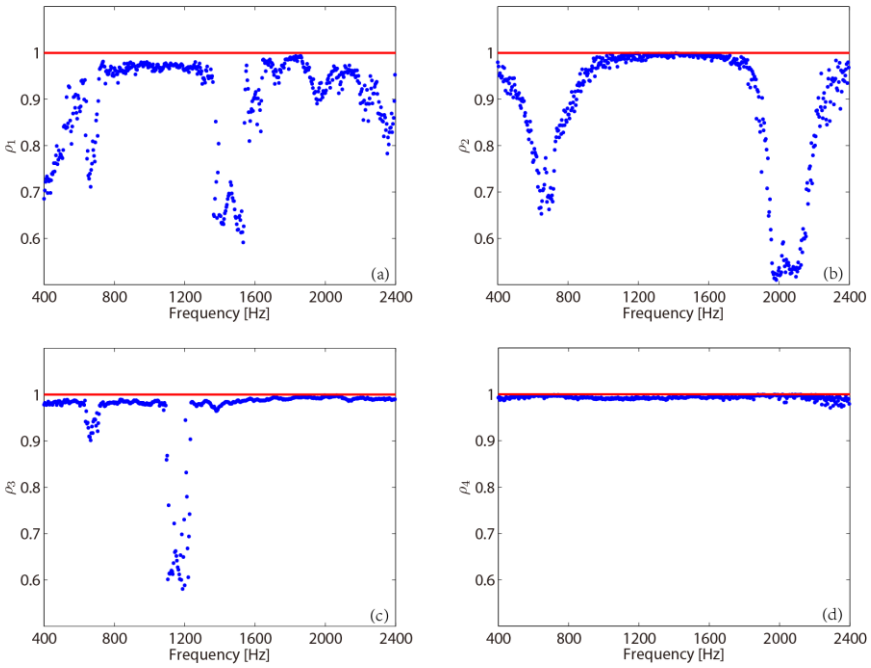

Fig. 6.21 Spatial correlation spectra between the separated and actual sources, $\rho_i(\omega)$, i = 1,…,4, from joint approximate diagonalization (blue point) compared to actual sources (red solid line), in the near-field.

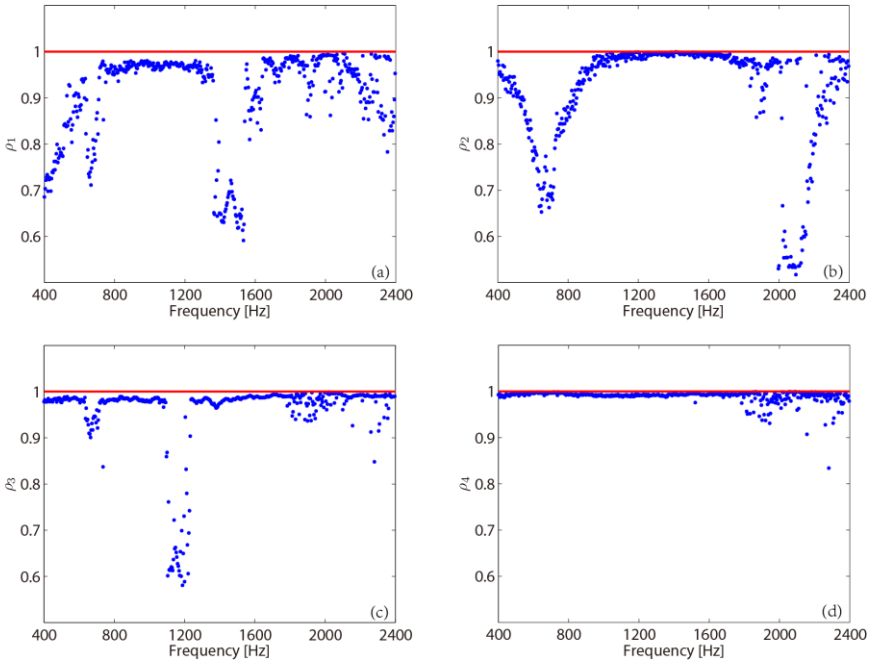

Fig. 6.22 Spatial correlation spectra between the separated and actual sources, $\rho_i(\omega)$, i = 1,…,4, from the conjugate gradient method (blue point) compared to actual sources (red solid line), in the near-field.

At 833 Hz, the spatial correlations coefficients between the actual and the separated sources, from single statistical decorrelation, joint statistical and spatial decorrelation with JAD and with CG, respectively, are listed in Tab. 6.2. There is no difference between the separated results from the spatial decorrelation with JAD and with CG, respectively.

	s_1	s_2	s_3	s_4
SD	0.82	0.88	0.86	1
2SD-JAD	0.97	0.91	0.99	0.99
2SD-CG	0.97	0.91	0.99	1

Tab. 6.2 Spatial correlation coefficients between the actual and the separated sources, from single statistical decorrelation (SD), joint statistical and spatial decorrelation (2SD) with JAD and with CG, respectively, in the near-field at 833 Hz.

2) **Source separation in the far-field**

The separated sources from SD, 2SD with JAD and CG, and the actual sources are all displayed in Figs. 6.23-6.26, respectively. After combing statistical and spatial decorrelation, the separated sources, no matter from JAD or from CG, coincide perfectly with the actual sources.

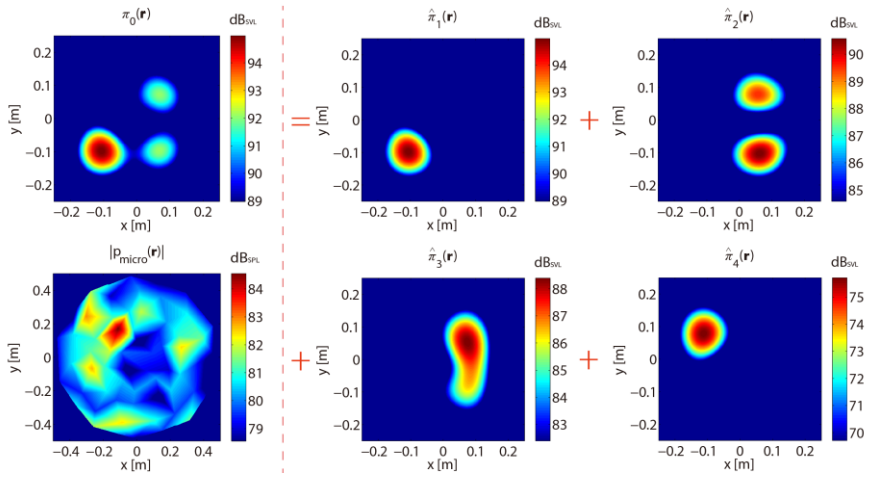

Fig. 6.23 Separated sources, $\hat{\pi}_i(\mathbf{r})$, $i = 1,\ldots,4$, from single statistical decorrelation, in the far-field at 2437 Hz.

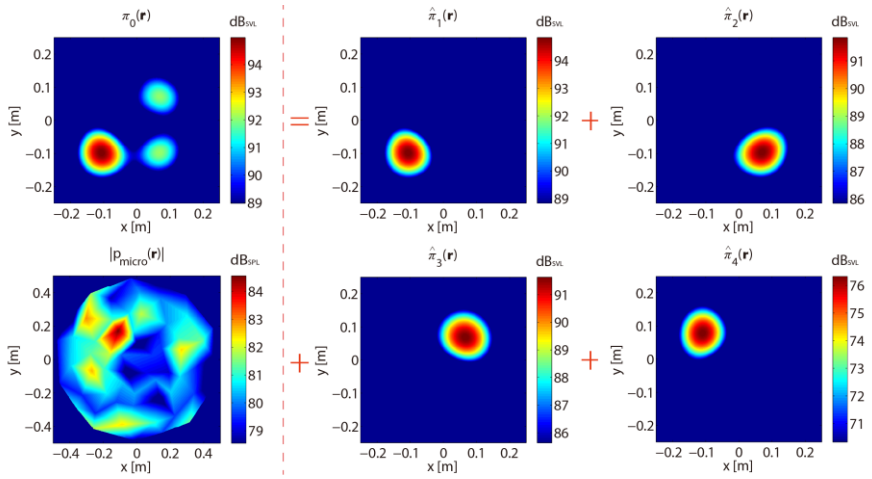

Fig. 6.24 Separated sources, $\hat{\pi}_i(\mathbf{r})$, $i = 1,\ldots,4$, from statistical and spatial decorrelation with joint approximate diagonalization, in the far-field at 2437 Hz.

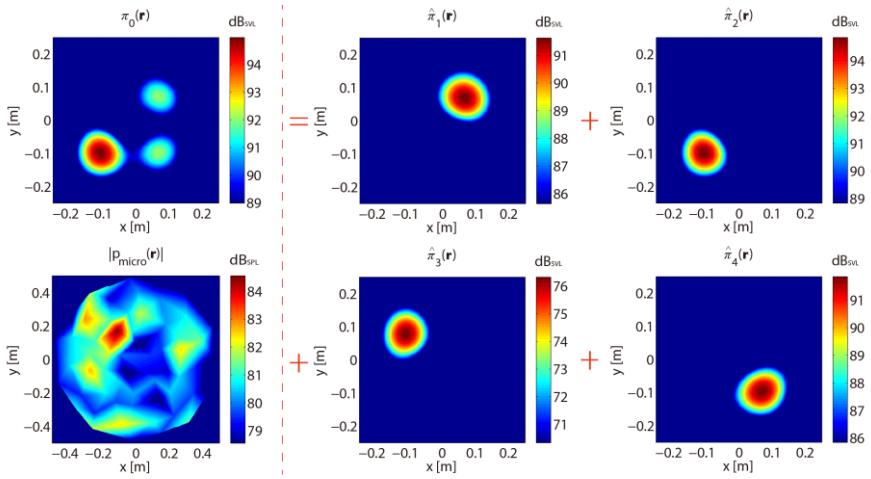

Fig. 6.25 Separated sources, $\hat{\pi}_i(\mathbf{r})$, $i = 1,\ldots,4$, from statistical and spatial decorrelation with conjugate gradient method, in the far-field at 2437 Hz.

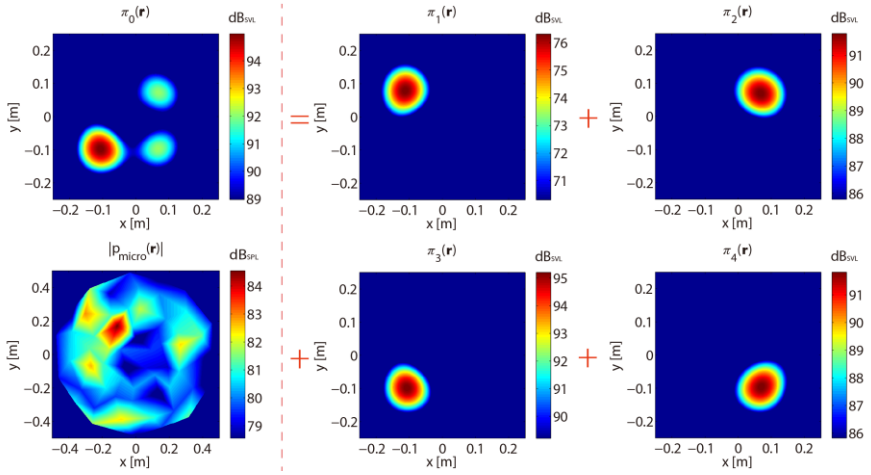

Fig. 6.26 Actual sources, $\pi_i(\mathbf{r})$, $i = 1,\ldots,4$, from individual measurements, in the far-field at 2437 Hz.

The quadratic velocity and spatial correlation spectra of the separated sources are shown in Figs. 6.27-6.30. Just as in the near-field case, the quadratic velocity spectra match closely to the references except maybe near the zeros and in the low frequency interval [1600 2400] Hz, as depicted in Figs. 6.27-6.28. The good performance of the algorithms, JAD and CG, is also checked by the spatial correlation spectra (see Figs. 6.29-6.30).

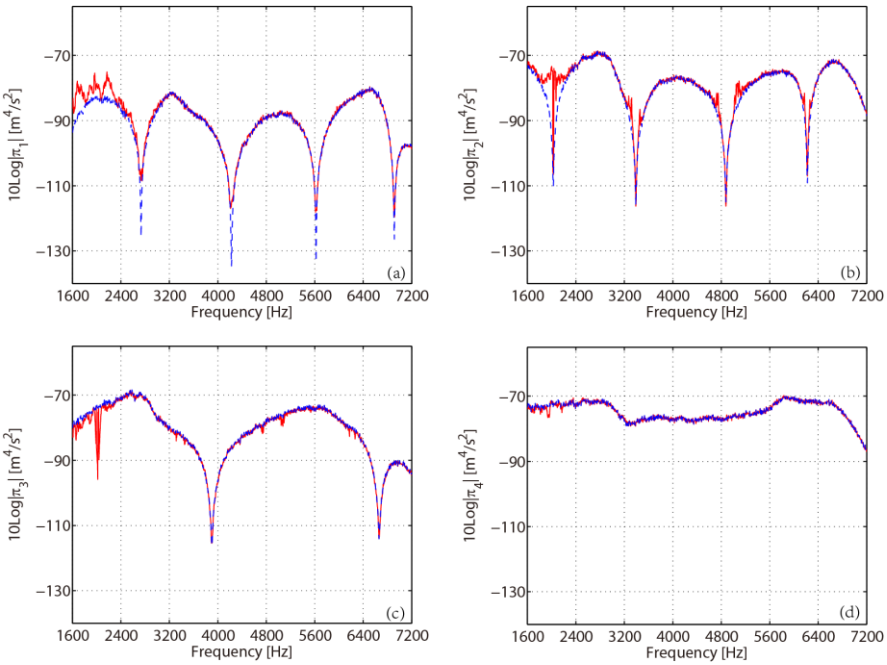

Fig. 6.27 Quadratic velocity spectra of the separated sources, $\hat{\pi}_i(\omega)$, $i = 1,\ldots,4$, from statistical and spatial decorrelation with joint approximate digonalization (red solid line) compared to actual sources (blue dashed line), in the far-field.

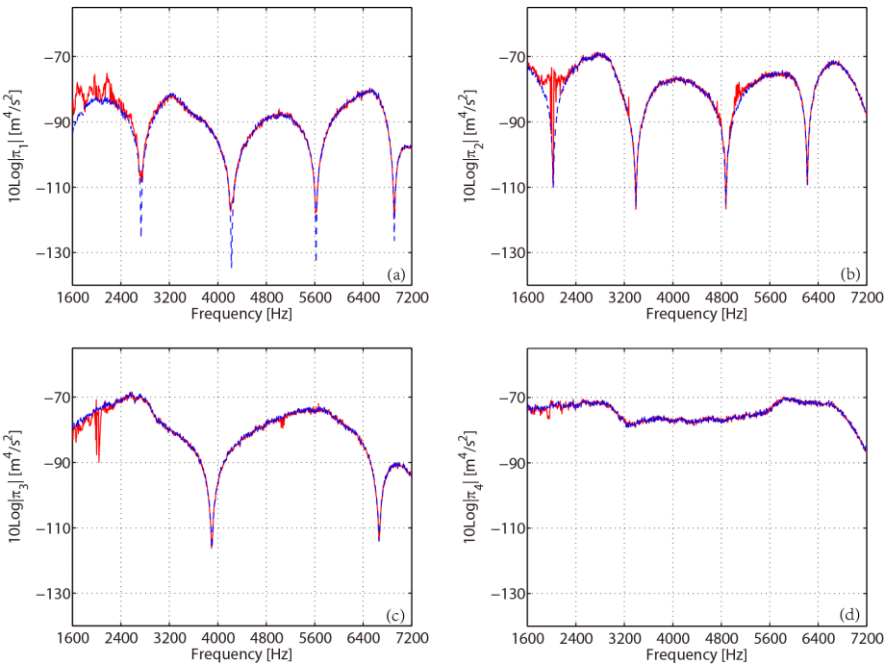

Fig. 6.28 Quadratic velocity spectra of the separated sources, $\hat{\pi}_i(\omega)$, $i = 1,\ldots,4$, from statistical and spatial decorrelation with conjugate gradient method (red solid line) compared

to actual sources (blue dashed line), in the far-field.

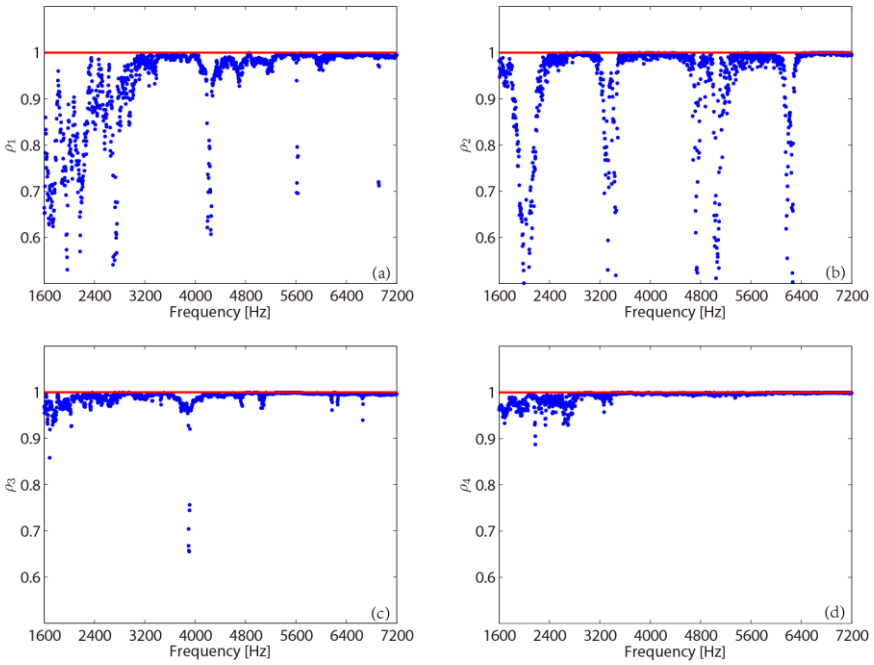

Fig. 6.29 Spatial correlation spectra between the separated and actual sources, $\rho_i(\omega)$, i = 1,…,4, from joint approximate diagonalization (blue point) compared to actual sources (red solid line), in the far-field.

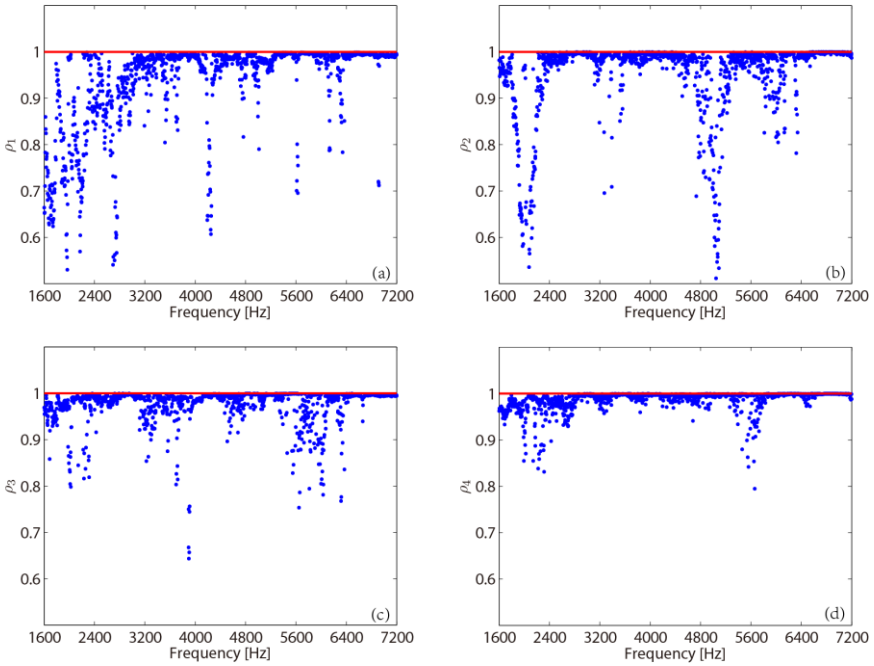

Fig. 6.30 Spatial correlation spectra between the separated and actual sources, $\rho_i(\omega)$, i = 1,…,4, from the conjugate gradient method (blue point) compared to actual sources (red solid line), in the far-field.

At 2437 Hz, the spatial correlations coefficients between the actual and the separated sources, from single statistical decorrelation, joint statistical and spatial decorrelation with JAD and with CG, respectively, are listed in Tab. 6.3. The difference between the results is insignificant from joint statistical and spatial decorrelation with JAD and CG. The minimum spatial correlation is still equal to 0.97.

	s_1	s_2	s_3	s_4
SD	1	0.86	1	0.84
2SD-JAD	0.97	1	0.99	0.98
2SD-CG	0.99	1	1	1

Tab. 6.3 Spatial correlation coefficients between the actual and the separated sources, from single statistical decorrelation (SD), joint statistical and spatial decorrelation (2SD) with JAD and with CG, respectively, in the far-field at 2437 Hz.

In acoustical holography, the spatial orthogonality of actual sources has no reason to be preserved in the whole work frequency band. Due to small wave length – i.e. high spatial resolution at high frequency, the sizes of reconstructed sources are more similar to the actual ones and thus the spatial orthogonality is well inherited by the reconstructed sources. Unfortunately, the sizes of reconstructed sources become much larger than their actual counterparts at low and medium frequencies because of the long wave length (i.e. the low spatial resolution). In such a case, the reconstructed sources overlap to each other in space. Therefore, SD can provide pretty accurate results for BSS at high frequency (as explained in Subsection 3.1.2), but degenerates at low and medium frequencies, because the matrix **V** is no longer the normalized permutation matrix but a general unitary one at low and medium frequencies.

6.1.4 Separation from the principle of least spatial complexity

The experimental parameters are all set as listed in Subsections 6.1.2-6.1.3 except for the radius of the aperture function which is set to $R = 21$cm.

1) **Source separation in the near-field**

First, virtual sources are computed from the eigenvalue decomposition of the correlation matrix $\mathbf{C}_{\hat{\mathbf{c}}\hat{\mathbf{c}}^H}$ and displayed on the right side of Fig. 6.31; this shows that the separated sources from SD depart significantly from the actual ones, which is similar to that in Fig. 6.15 (ignoring the errors from backpropagation by applying two distinct aperture functions – i.e. $R = 21$ cm and 35 cm, respectively, the two groups of virtual sources (see Figs. 6.31 and 6.15) are almost the same in terms of locations and amplitude of the hottest points). The latter is then estimated by using CLSV and CLSE. The separation results are displayed in Figs. 6.32-6.33, respectively. The 4 actual sources, in the near-field at 833 Hz, are displayed in Fig. 6.34.

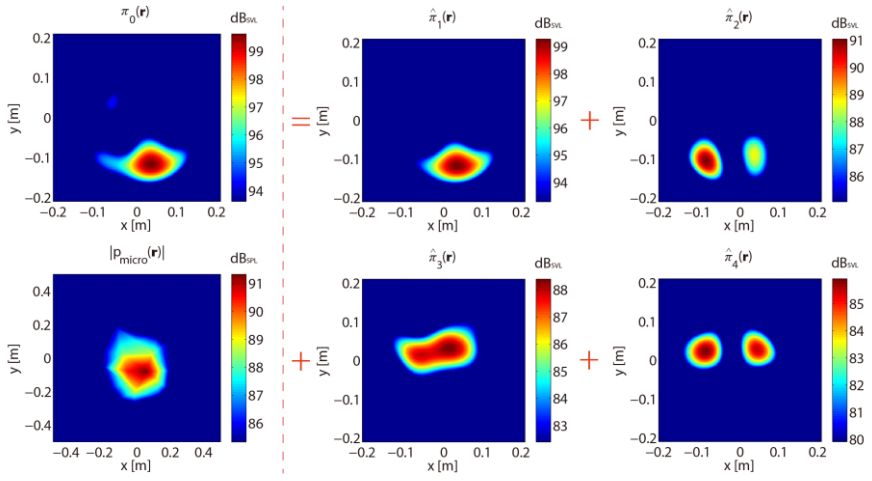

Fig. 6.31 Separated sources, $\hat{\pi}_i(\mathbf{r},\omega)$, $i = 1,\ldots,4$, from single statistical decorrelation, in the near-field at 833 Hz.

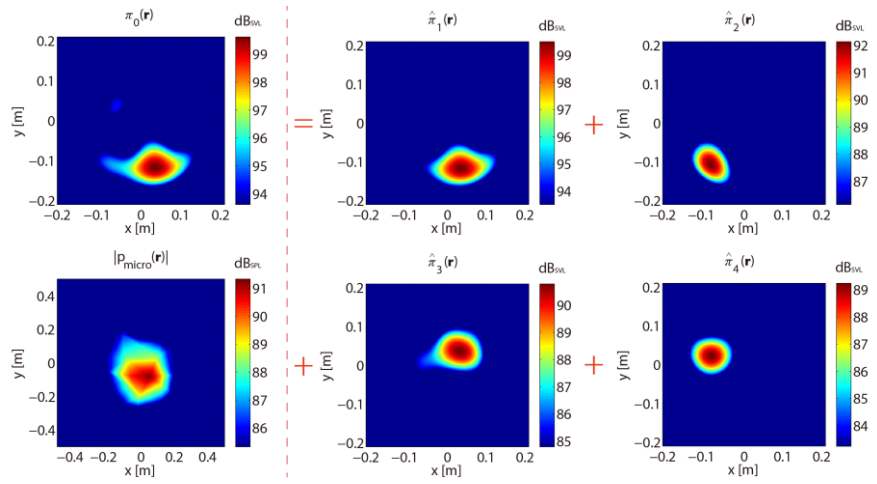

Fig. 6.32 Separated sources, $\hat{\pi}_i(\mathbf{r},\omega)$, $i = 1,\ldots,4$, from the criterion of least spatial variance, in the near-field at 833 Hz.

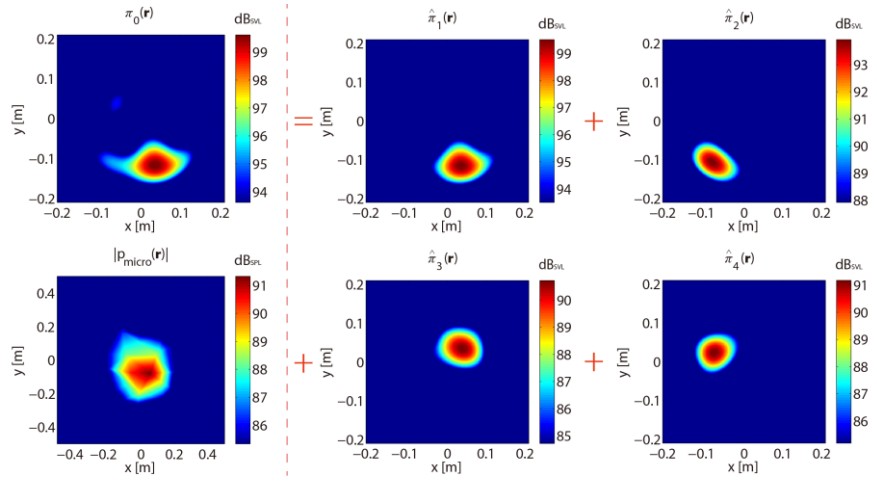

Fig. 6.33 Separated sources, $\hat{\pi}_i(\mathbf{r},\omega)$, $i = 1,\ldots,4$, from the criterion of least spatial entropy, in the near-field at 833 Hz.

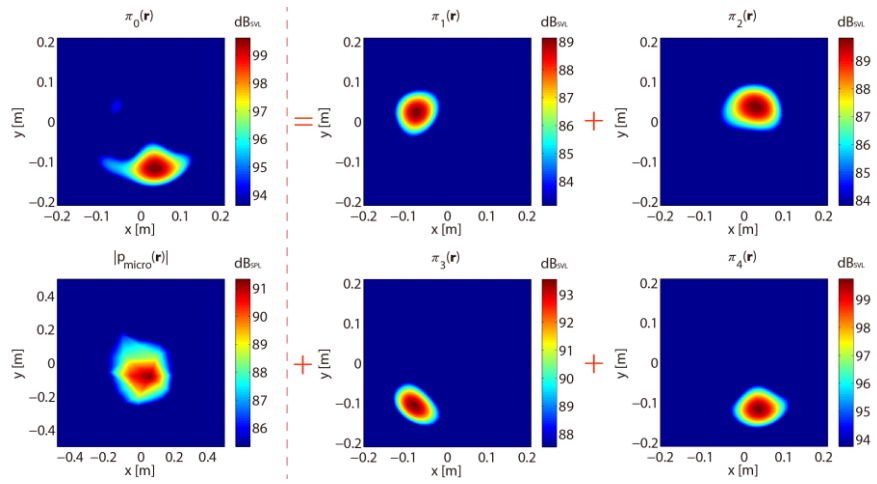

Fig. 6.34 Actual sources, $\pi_i(\mathbf{r},\omega)$, i = 1,…,4, from individual measurements, in the near-field at 833 Hz.

Comparing Figs. 6.31-6.34 with Figs. 6.15-6.18, the size of aperture function has insignificant effect on BSS in the near-field.

It is seen that the separation results from the two criteria return accurate localization and quantification of the sources, as compared to the actual ones (up to an arbitrary labeling). On this example CLSV seems to be slightly more accurate than CLSE in estimating amplitude, but slightly inferior as for the spatial localization; indeed, the separated source $\hat{\pi}_2(\mathbf{r},\omega)$ from CLSV has a small protrusion on its left part. More generally, CLSE is expected to be more robust to the separation of weak sources owing to the logarithmic operator which reduces the dynamic range between the lowest and the loudest sources. On the contrary, the squaring operation inherent to CLSV will increase the dynamic range and possibly jeopardize the recovery of sources with relatively small magnitude. However, it is observed that CLSV has a much faster convergence speed than CLSE in the optimization procedure – it is about three times faster.

Noteworthy also is that the spatial distributions of some separated sources (e.g. $\hat{\pi}_1(\mathbf{r},\omega)$ and $\hat{\pi}_2(\mathbf{r},\omega)$) are slightly more compact than the actual sources, as a result of enforcing this property in the separation; this small bias is however not dramatic and may be even seen as an advantage since spatial resolution is thus improved.

Eventually, the separated results from the two spatial criteria are compared with those from a classic method – SOBI [[48], [71]]. SOBI was applied in a similar fashion as in [69] so as to jointly diagonalise the correlation matrix $\mathbf{C}_{\hat{\mathbf{c}}\hat{\mathbf{c}}^H}$ in a narrow frequency band centered at 833 Hz with bandwidth 1.75 Hz (i.e. 7 frequency bins). The separated sources are displayed in Fig. 6.35. SOBI could find only two distinct sources, $\hat{\pi}_2(\mathbf{r},\omega)$ and $\hat{\pi}_3(\mathbf{r},\omega)$, yet with severe disturbance and strong coupling in space. Similar comparisons carried on in other frequency ranges generally confirmed that SOBI could hardly reach the same separation performance as LCSV and CLSE because it does not account for spatial information, as proposed in this dissertation.

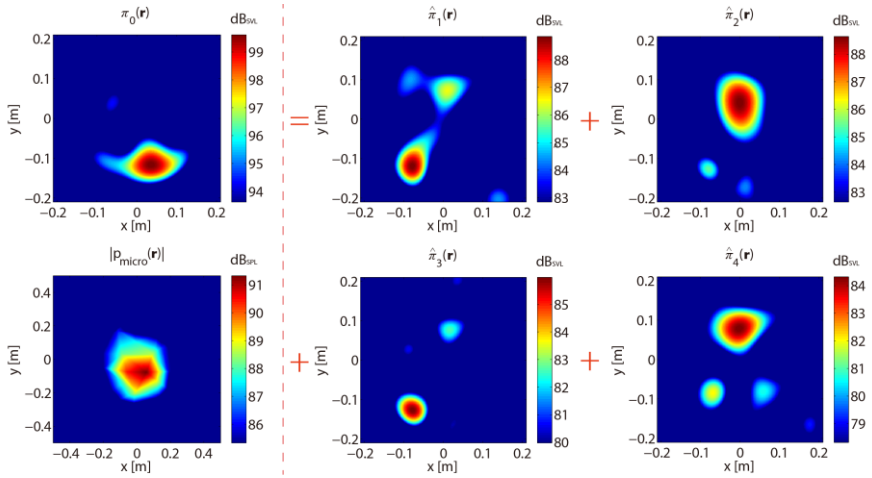

Fig. 6.35 Separated sources, $\hat{\pi}_i(\mathbf{r},\omega)$, i = 1,…,4, from SOBI, in the near-field at 833 Hz.

Next, the separation results are displayed in terms of quadratic velocity spectra in the whole available frequency band for CLSV in Fig. 6.36 and for CLSE in Fig. 6.37. Note that the lower frequency bound is determined by the size of the microphone array and the upper one by the minimum spacing between two microphones.

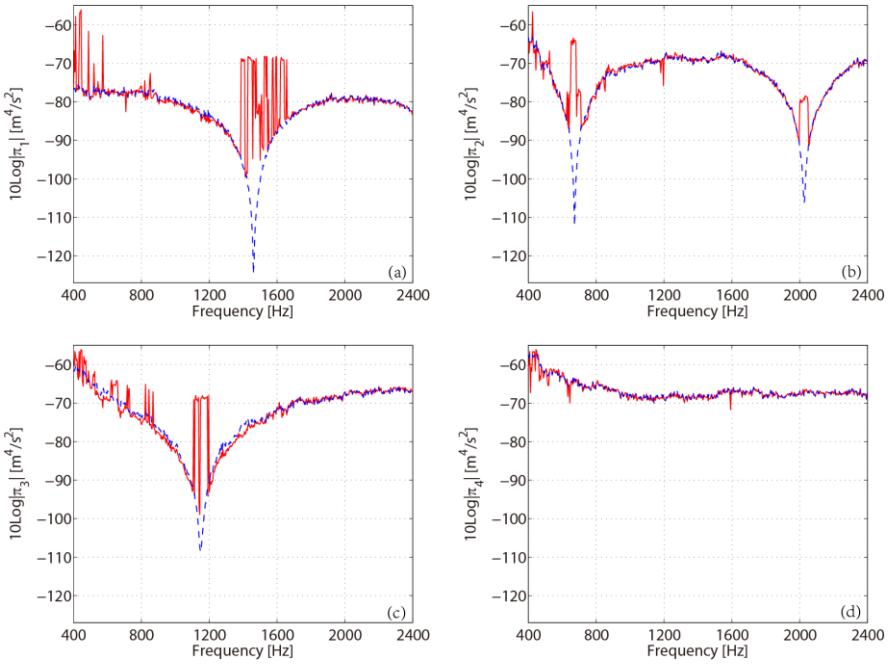

Fig. 6.36 Quadratic velocity spectra of the separated sources, $\hat{\pi}_i(\omega)$, $i = 1,\ldots,4$, from the criterion of least spatial variance (red solid line) compared to actual sources (blue dashed line), in the near-field.

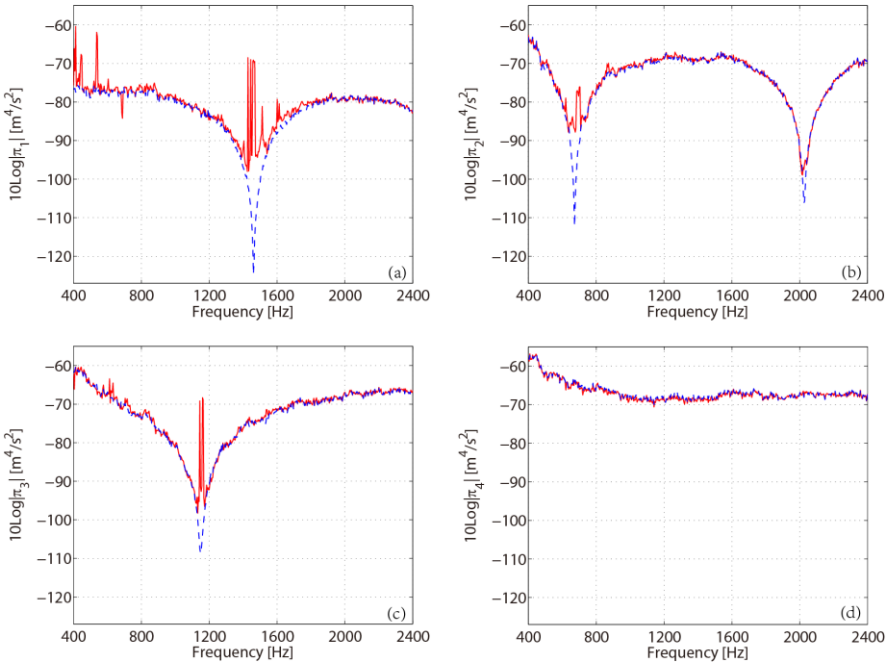

Fig. 6.37 Quadratic velocity spectra of the separated sources, $\hat{\pi}_i(\omega)$, $i = 1,\ldots,4$, from the criterion of least spatial entropy (red solid line) compared to actual sources (blue dashed line),

in the near-field.

Separation results are satisfactory for most frequencies, except at some points in low frequencies (mainly in interval [400, 600] Hz) and near zeros of source spectra. This may be explained as follows. In the near-field, the proposed method is inherently limited by the inability of the backpropagation process to estimate the source distributions below a certain lower frequency (that depends on the array size and signal-to-noise ratio). In this respect, CLSE seems more robust to CLSV, probably because it is less sensitive to the presence of additive noise for the reason invoked above. Moreover, at those frequencies where the energy of the source is so small (zeros in the spectra) that virtually only noise is measured, it makes sense that source separation is doomed to failure. For instance, in the frequency band [1300, 1500] Hz enclosing a zero in the quadratic velocity spectrum of the actual source s_1 both criteria fail to separate the source due to its very small magnitude (see Fig. 6.36(a) and Fig. 6.37(a)). However, it is seen that the sources separated from CLSE have in general spectra closer to the actual ones than for CLSV. Again, this illustrates the higher robustness of CLSE against additive noise.

The spatial correlation spectra of the separation results in the whole working frequency band are presented for CLSV in Fig. 6.38 and for CLSE in Fig. 6.39. The spatial correlation spectra demonstrate again that the two criteria blindly separate the incoherent sound sources very well in the whole frequency band, including the zeros, e.g. the frequency band [1300, 1500] Hz, and the low frequency band [400, 600] Hz. Comparing Figs. 6.38 and 6.39, CLSE works better than CLSV not only in the regions with zeros but also at low frequencies; from the point of view of the spatial correlation spectrum, CLSE is also more robust to amplitude differences among the sound sources.

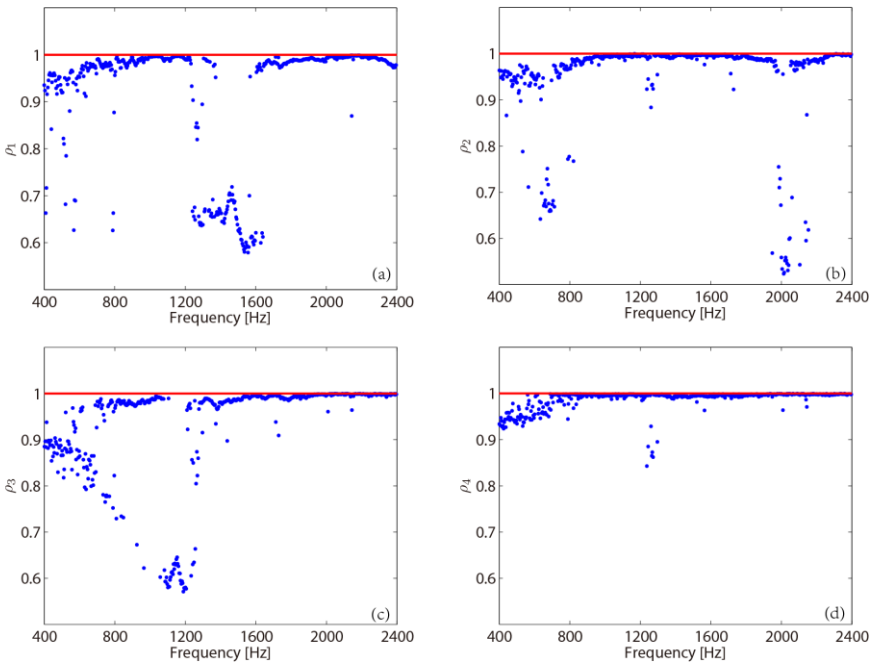

Fig. 6.38 Spatial correlation spectra between the separated and actual sources, $\rho_i(\omega)$, i = 1,…,4, from the criterion of least spatial variance, in the near-field.

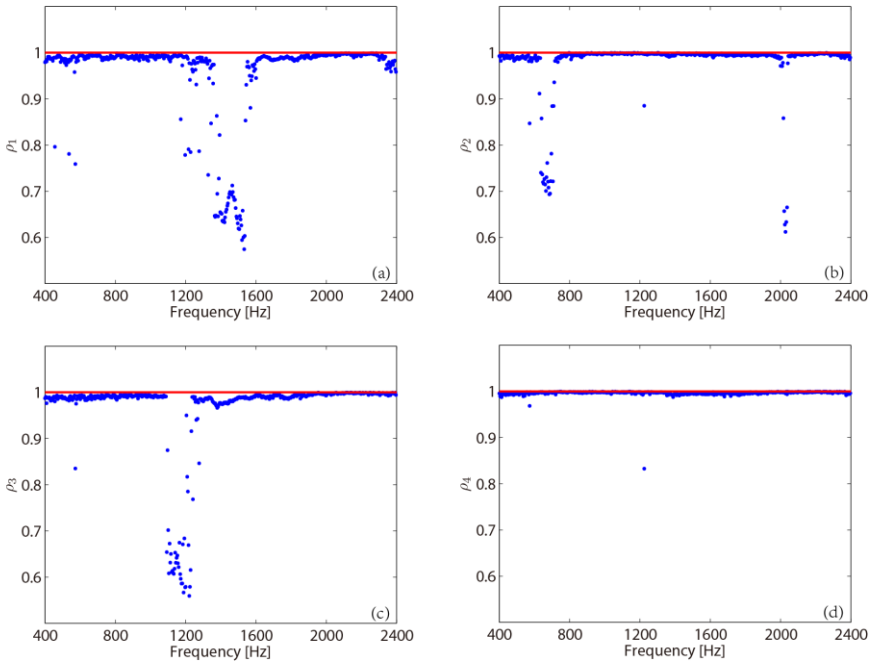

Fig. 6.39 Spatial correlation spectra between the separated and actual sources, $\rho_i(\omega)$, i = 1,…,4, from the criterion of least spatial entropy, in the near-field.

2) **Source separation in the far-field**

The separated sources from enforcing mutual incoherence, from CLSV and CLSE, and the actual sources are all displayed in Figs. 6.40-6.43, respectively. It is seen that the reconstructed source distribution $\pi_0(\mathbf{r},\omega)$ essentially consists in one dominant source plus two or three very small ones. The sources separated from enforcing only mutual incoherence look compact enough, but unfortunately the second and third ones do not match with the actual sources. On the other hand, CLSV and CLSE return excellent results. The sources are perfectly separated, even the smallest one which stands 20dB below the dominant one.

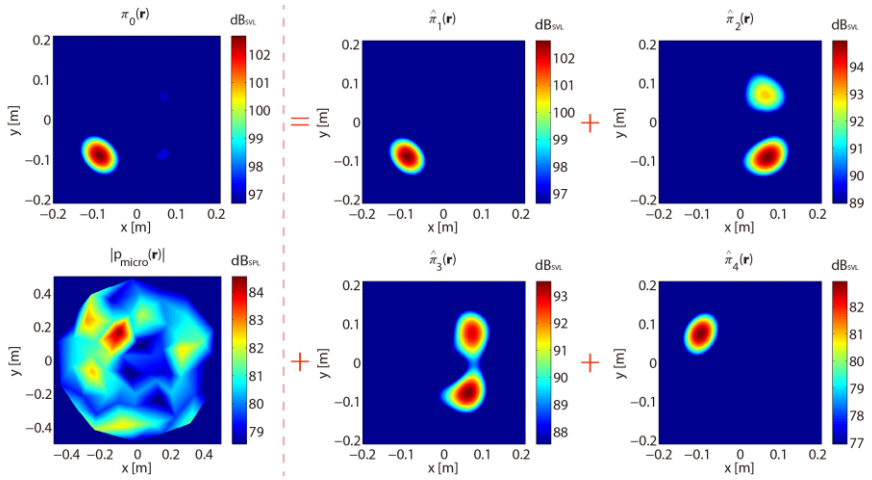

Fig. 6.40 Separated sources, $\hat{\pi}_i(\mathbf{r},\omega)$, $i = 1,\ldots,4$, from single statistical decorrelation, in the far-field at 2437 Hz.

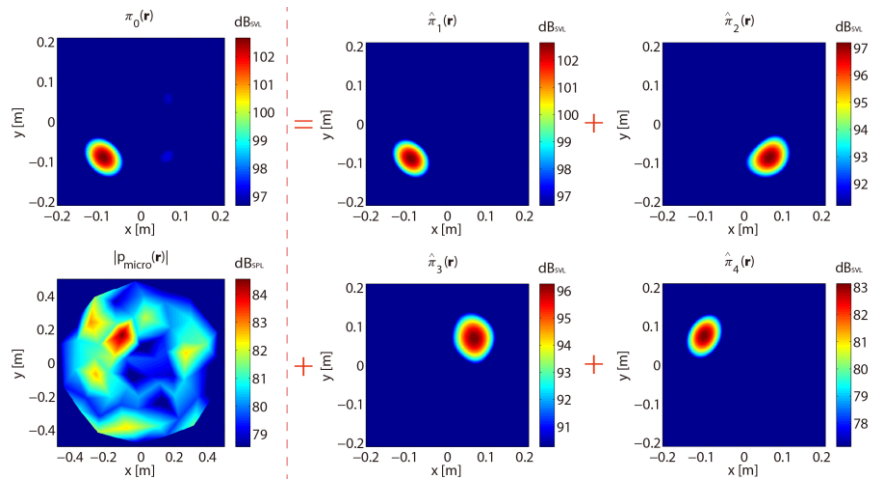

Fig. 6.41 Separated sources, $\hat{\pi}_i(\mathbf{r},\omega)$, $i = 1,\ldots,4$, from the criterion of least spatial variance, in the far-field at 2437 Hz.

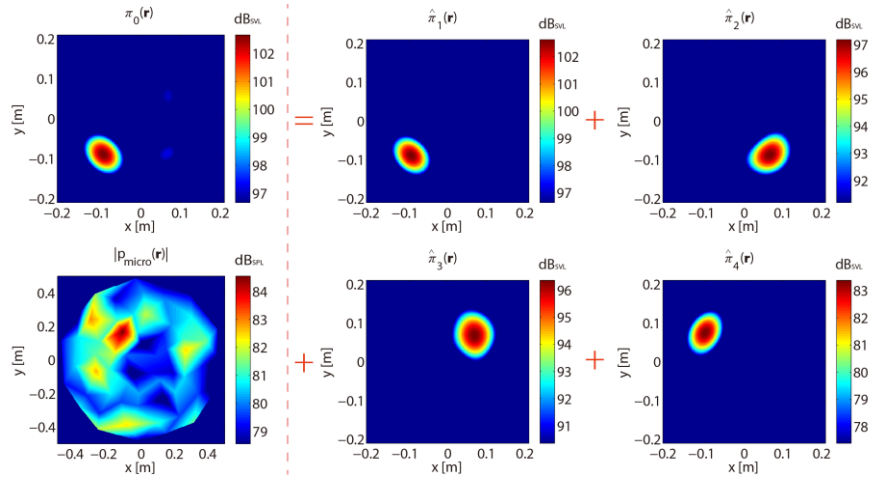

Fig. 6.42 Separated sources, $\hat{\pi}_i(\mathbf{r},\omega)$, $i = 1,\ldots,4$, from the criterion of least spatial entropy, in the far-field at 2437 Hz.

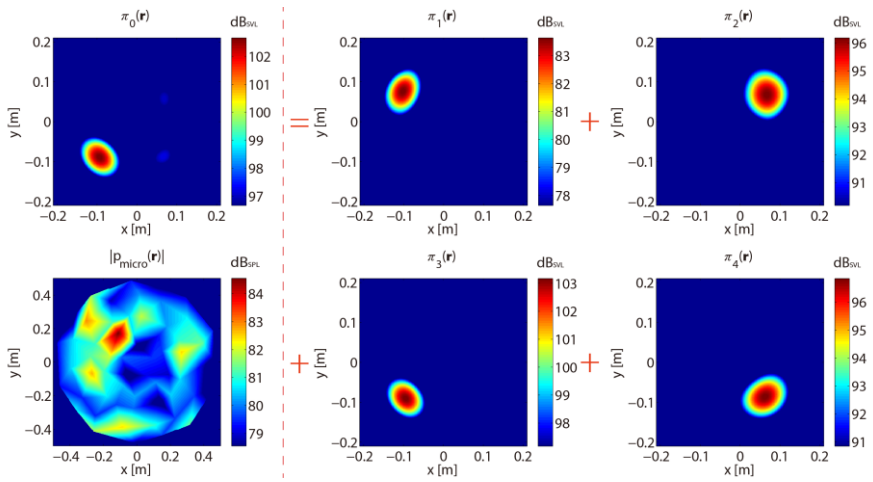

Fig. 6.43 Actual sources, $\pi_i(\mathbf{r},\omega)$, $i = 1,\ldots,4$, from individual measurements, in the far-field at 2437 Hz.

The separation results returned by SOBI (with similar settings as before) are illustrated in Fig. 6.44. In this case, the separated sources look extremely scattered and can hardly be identified with the actual sources.

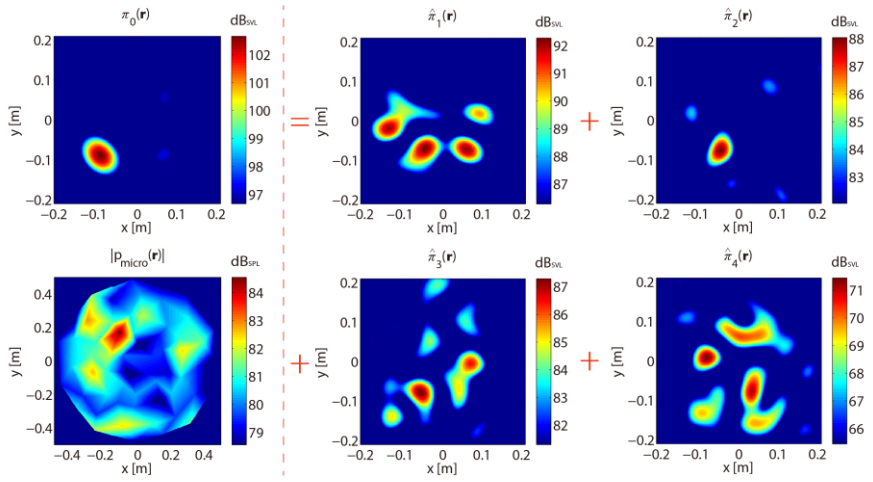

Fig. 6.44 Separated sources, $\hat{\pi}_i(\mathbf{r},\omega)$, $i = 1,\ldots,4$, from SOBI, in the far-field at 2437 Hz.

The quadratic velocity spectra of the separated sources are shown in Figs. 6.45-6.46, and their spatial correlation spectra in Figs. 6.47-6.48. Note that the frequency range could be extended much higher (up to 7200Hz) because the measurements are taken in the far-field. Again, CLSE evidences a better performance, in particular in its

ability to separate sources with very small levels. Therefore, CLSE is recommended as the first choice due to its robustness, provided the extra calculation cost it implies is affordable.

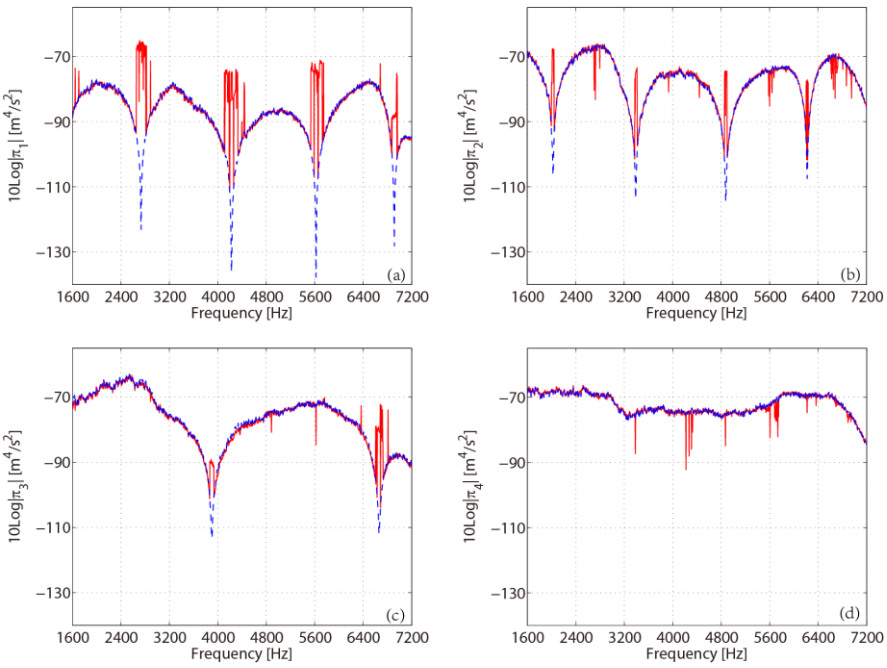

Fig. 6.45 Quadratic velocity spectra of the separated sources, $\hat{\pi}_i(\omega)$, i = 1,…,4, from the criterion of least spatial variance (red solid line) compared to actual sources (blue dashed line), in the far-field.

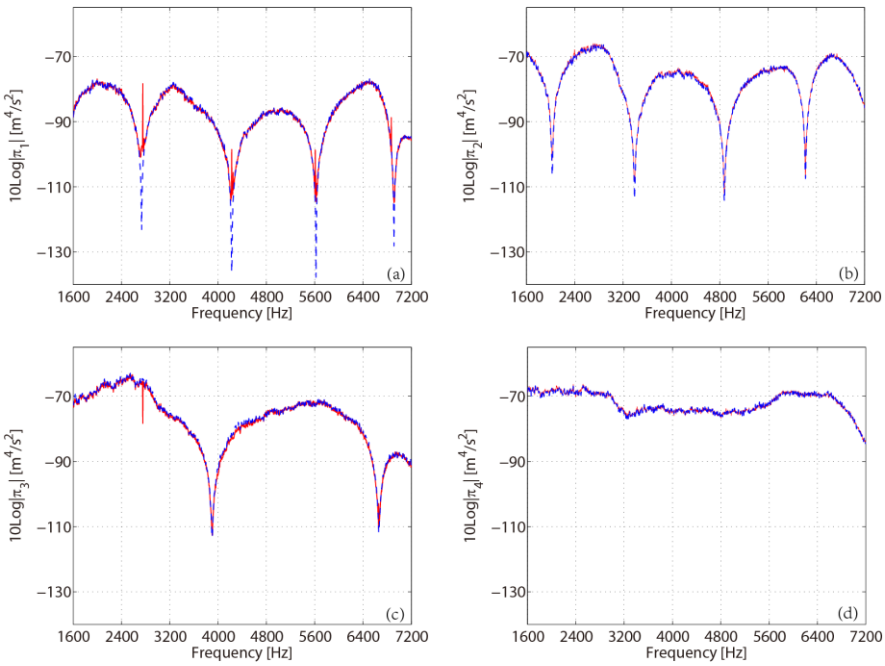

Fig. 6.46 Quadratic velocity spectra of the separated sources, $\hat{\pi}_i(\omega)$, i = 1,…,4, from the criterion of least spatial entropy (red solid line) compared to actual sources (blue dashed line), in the far-field.

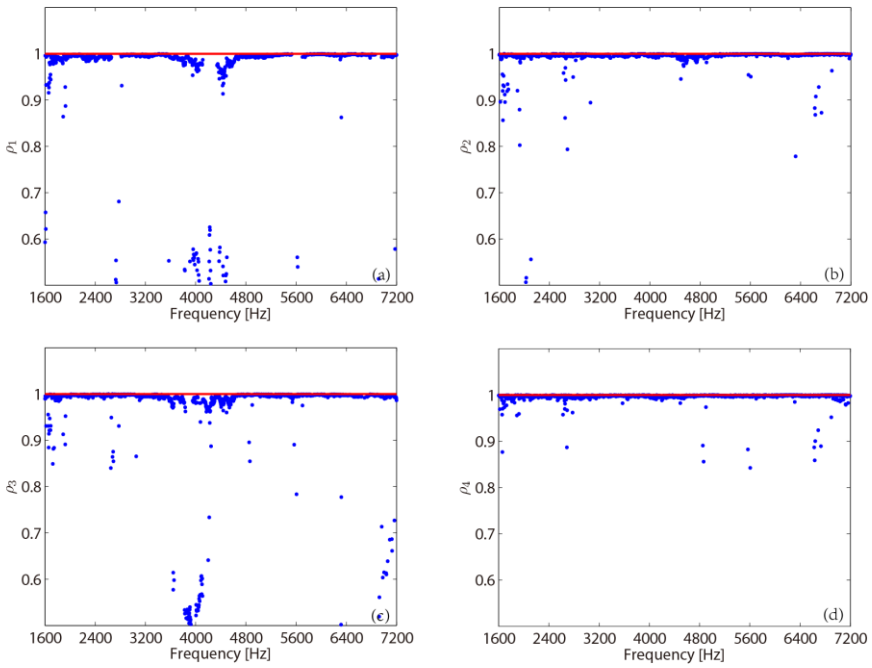

Fig. 6.47 Spatial correlation spectra between the separated and actual sources, $\rho_i(\omega)$, i = 1,…,4, from the criterion of least spatial variance, in the far-field.

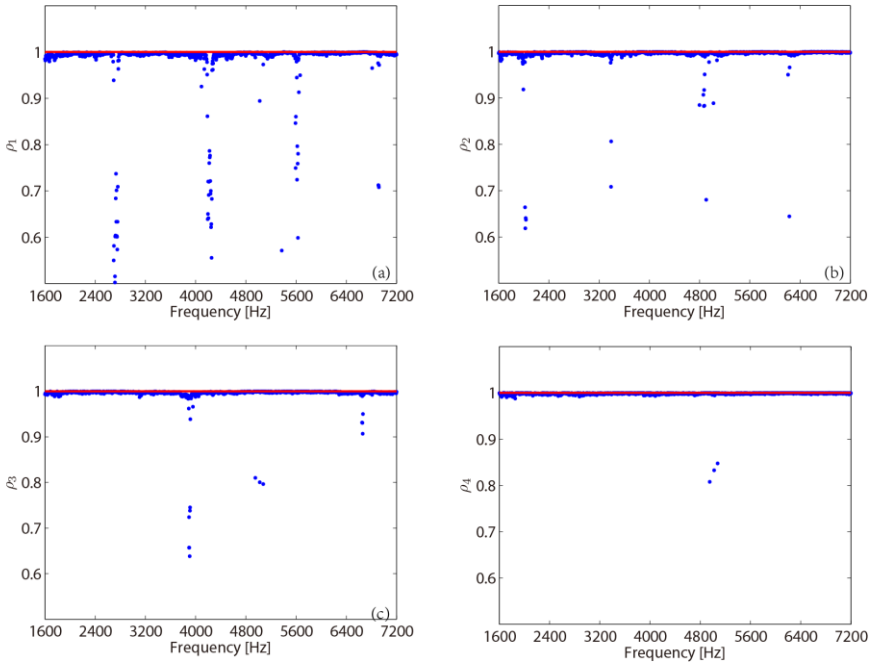

Fig. 6.48 Spatial correlation spectra between the separated and actual sources, $\rho_i(\omega)$, i = 1,…,4, from the criterion of least spatial entropy, in the far-field.

6.1.5 Separation from ranking eigen elements according to increasing spatial entropy

As seen in Figs. 6.19-6.22, 6.27-6.30, 6.36-6.39, and 6.45-6.48, the three criteria 2SD, CLSV and CLSE could hardly separate sources with very small levels, in particular near the zeros of the quadratic velocity spectra (although CLSE was found more robust to small sources). It is shown in this section that low sources can be found by ranking the eigen elements of the eigenspectrum with respect to the spatial information (measured by spatial variance or spatial entropy) of the virtual sources (see in Subsection 3.2.3). The scheme of ranking eigen elements according to increasing spatial information can be applied in CLSE and 2SD, but not in CLSV. As the square operation in CLSV, small sources might be ignored in the process of optimizing the total spatial variance of all sources.

The reason why the virtual sources are selected to rank the eigen elements of the eigenspectrum, not the ones from CLSE, is very simple and practical: the calculation of CLSE is unaffordable and might take several weeks, when the number of sources of interest, for example, arrives at 30 (see Fig. 6.49) and the spatial resolution of acoustic hologram is set to, for instance, 1 mm relative to the radius of the aperture function – R = 35 cm. In this subsection, CLSE is applied to blindly separate such small sources with the aforementioned scheme.

1) **Source separation in the near-field**

In the near-field, the radius of aperture function is set to R = 35 cm and the working frequency is selected as 1456 Hz where one source is much smaller than the other sources by more than 40 dB. The spatial entropies of the first 30 virtual sources are illustrated in Fig. 6.49. Obviously, the first three virtual sources plus the 15th one (marked by the red square) deserve special interest as their spatial distributions are the most compact among the first 30 largest virtual sources. Keeping only the corresponding eigen elements in matrices $\bar{\mathbf{D}}$ and $\bar{\mathbf{U}}$ used in CLSE, the very small source and the other three ones could be successfully separated as shown in Fig. 6.50. Note that the smallest separated source has amplitude 48 dB lower than the largest one, (see Fig. 6.51). To indicate the performance of the improved separation method, the quadratic velocity and spatial correlation spectra are displayed in Figs. 6.52-6.53, respectively. Comparing the two figures with Figs. 6.37 and 6.39, significant improvement is noted.

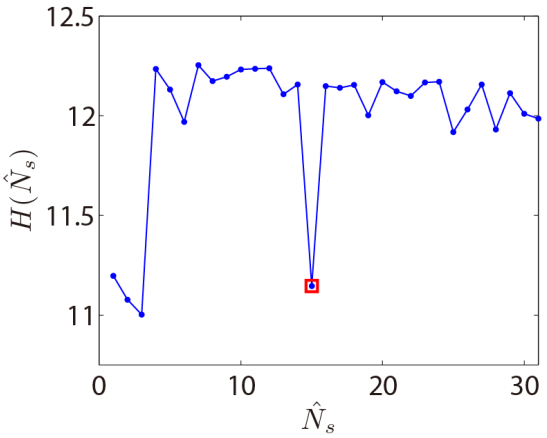

Fig. 6.49 Spatial entropy of the first 30 virtual sources in the near-field at 1456 Hz.

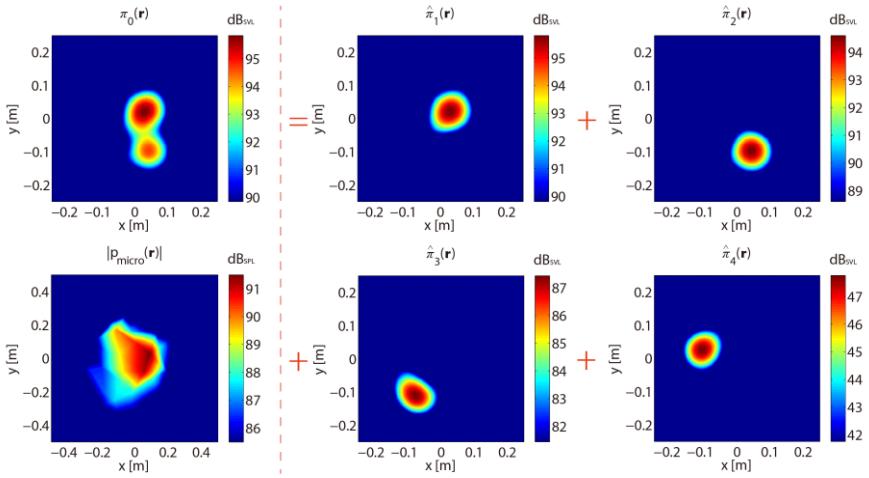

Fig. 6.50 Separated sources, $\hat{\pi}_i(\mathbf{r},\omega)$, i = 1,…,4, from the criterion of least spatial entropy with a new order, in the near-field at 1456 Hz.

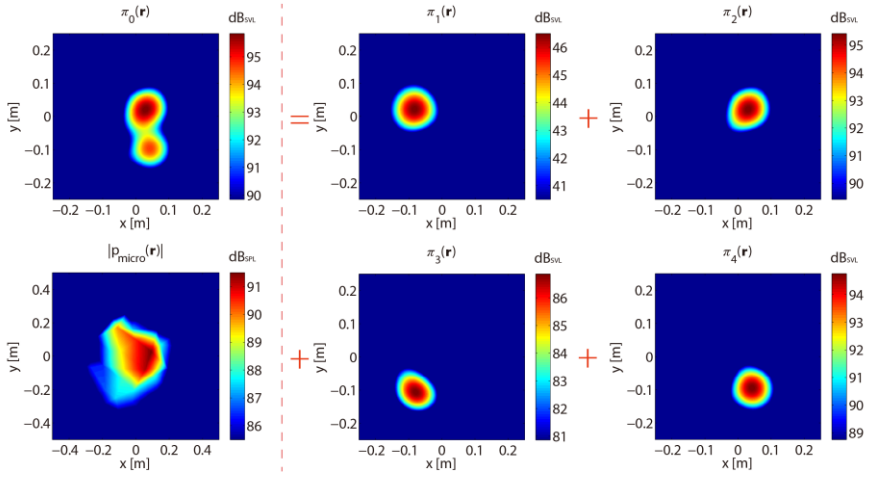

Fig. 6.51 Actual sources, $\pi_i(\mathbf{r},\omega)$, i = 1,…,4, from individual measurements, in the near-field

at 1456 Hz.

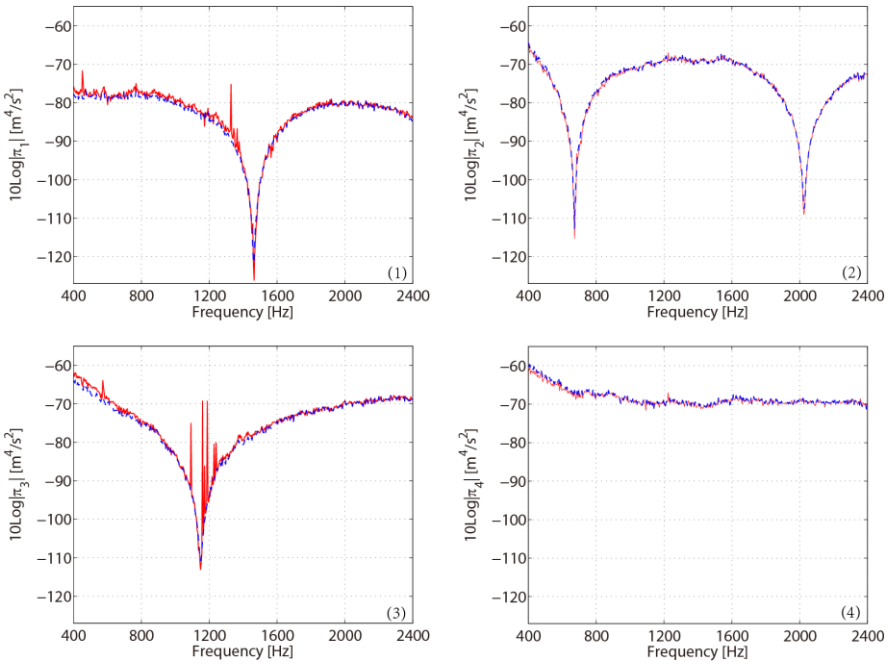

Fig. 6.52 Quadratic velocity spectra of the separated sources, $\hat{\pi}_i(\omega)$, i = 1,…,4, from the criterion of least spatial entropy (red solid line) compared to actual sources (blue dashed line), in the near-field.

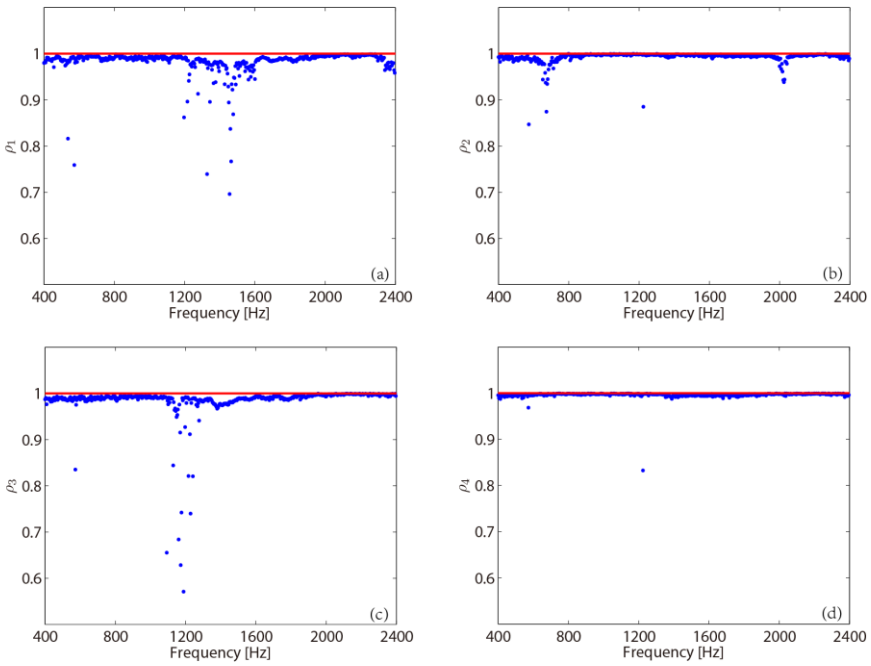

Fig. 6.53 Spatial correlation spectra between the separated and actual sources, $\rho_i(\omega)$, i = 1,…,4, from the criterion of least spatial entropy, in the near-field.

2) **Source separation in the far-field**

The spatial entropy of the first 30 virtual sources, the separated sources from the improved method, the corresponding references, and the quadratic velocity and spatial correlation spectra are depicted in Figs. 6.54-6.58, respectively. After selecting the eigen elements corresponding to the virtual sources with smallest spatial entropy (numbers 1, 2, 3 and 5 in Fig. 6.54), the smallest sound source could be correctly separated as shown in Fig. 6.54. The separation results are excellent in terms of amplitude and position as shown in Figs. 6.55-6.56. As seen in Figs. 6.57-6.58, the improved method could separate sound sources in the whole frequency band in the far-field with the help of CLSE, even at a few frequencies where the power of sound sources is similar to the background noise.

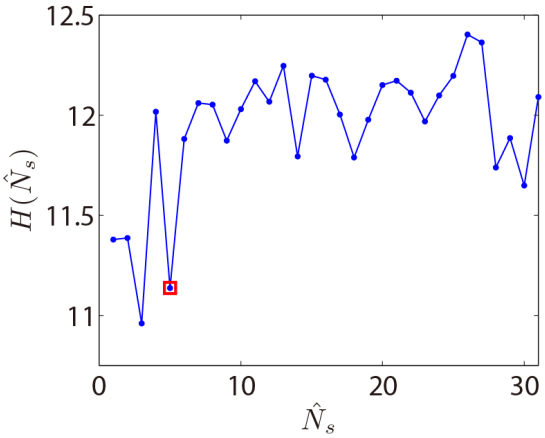

Fig. 6.54 Spatial entropy of the first 30 virtual sources in the far-field at 2025 Hz.

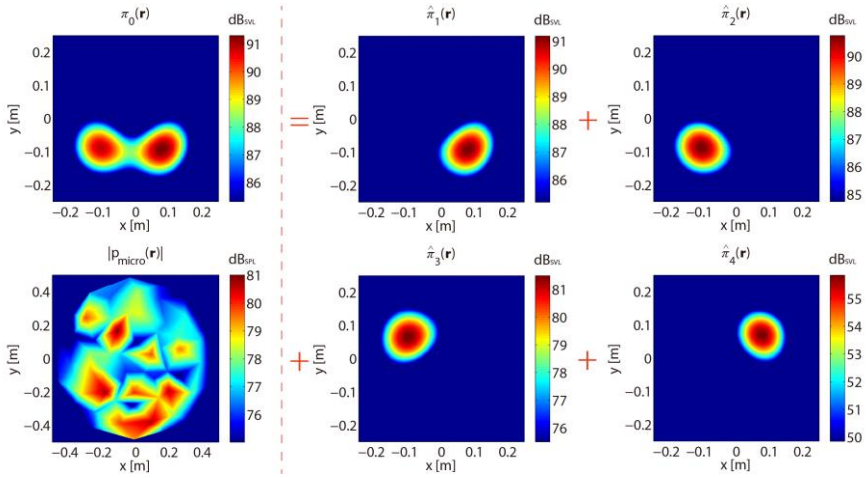

Fig. 6.55 Separated sources, $\hat{\pi}_i(\mathbf{r},\omega)$, i = 1,…,4, from the criterion of least spatial entropy with a new order, in the far-field at 2025 Hz.

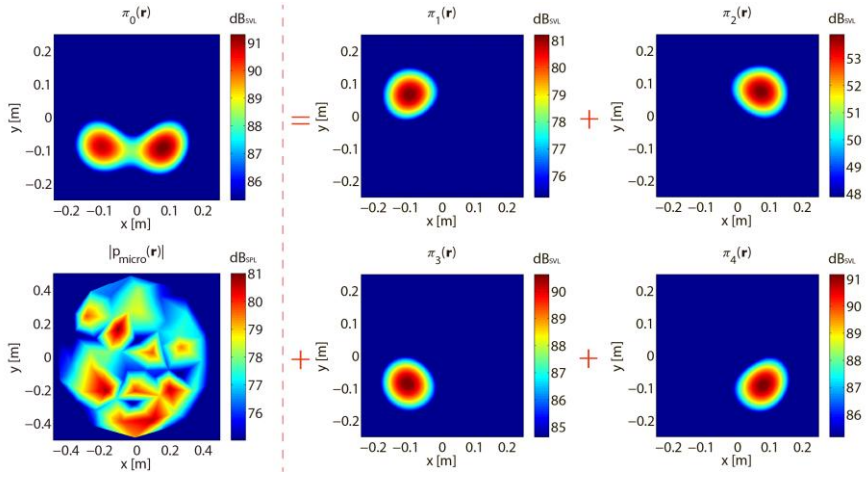

Fig. 6.56 Actual sources, $\pi_i(\mathbf{r},\omega)$, i = 1,…,4, from individual measurements, in the far-field

at 2025 Hz.

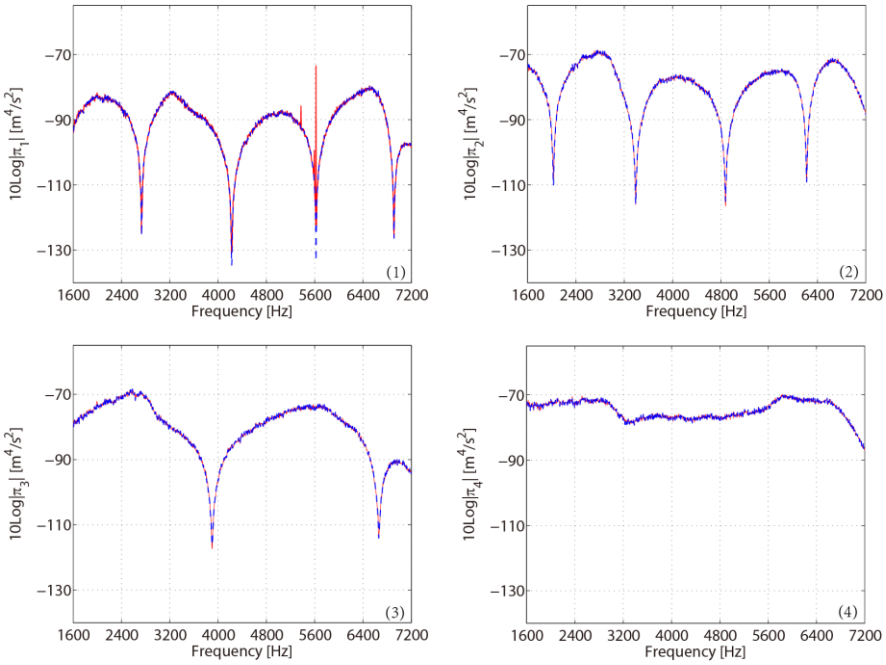

Fig. 6.57 Quadratic velocity spectra of the separated sources, $\hat{\pi}_i(\omega)$, i = 1,…,4, from the criterion of least spatial entropy (red solid line) compared to actual sources (blue dashed line), in the far-field.

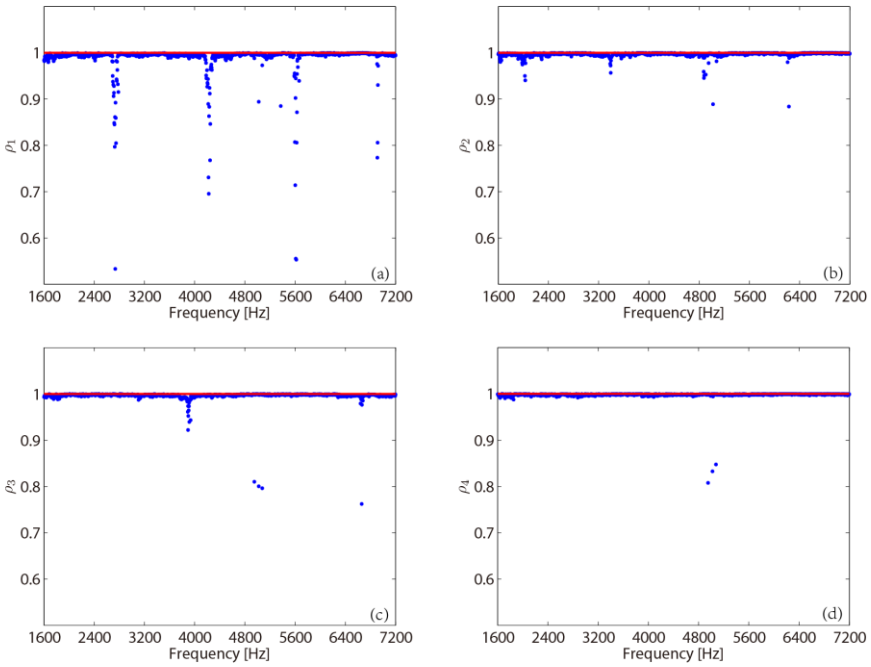

Fig. 6.58 Spatial correlation spectra between the separated and actual sources, $\rho_i(\omega)$, i = 1,…,4, from the criterion of least spatial entropy, in the far-field.

All experimental parameters and their values involved in Section 6.1 are listed in Tab. 6.4.

Z [cm]	10	50	100	200		
D [cm]	12	18	24			
f [Hz]	833	1456	1800	2025	2437	3600
R [cm]	21	35				

Tab. 6.4 Experimental values for all parameters in the Section 6.1.

6.2 Numerical experiments: simulation of the separation of multipoles

As demonstrated by the results of Subsection 6.1.4, the two proposed criteria have excellent performance for separating compact sources such as monopoles. As a matter

of fact, the property of compactness is the only requirement, and more complex sources than monopoles could just as well be separated. This is demonstrated by means of numerical simulations in this subsection, and by an industrial case in the next subsection.

Three superimposed sound sources are simulated: one monopole, one dipole, and one lateral quadrupole. The amplitudes of all sources are set identical. The microphone array is the same as in the previous experiments (slide wheel with 60 microphones). The distance Z between the array and the sound sources is 50 cm and the working frequency is 1 kHz. The SNR is set to 40 dB. Before separation, the source number N_s is determined from the information theoretic criteria (AIC and MDL), the eigenvalue spectrum, the cumulative power distribution, and the entropic L-curve displayed in Figs. 6.59-6.60, respectively. The five methods all return the correct result, $\hat{N}_s = 3$ (this is found at the inflexion point of AIC and MDL, not at their minimum).

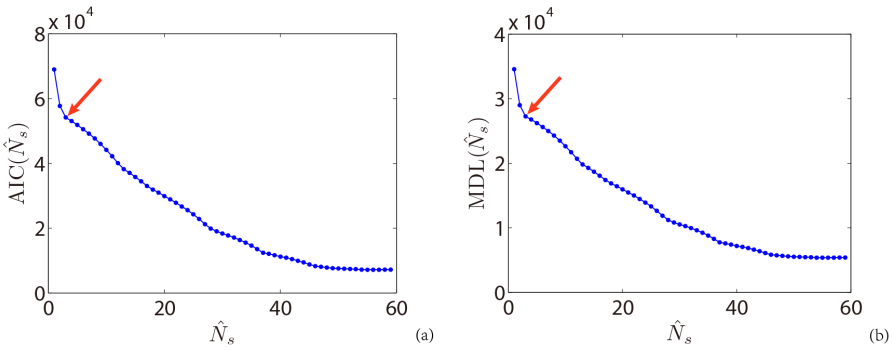

Fig. 6.59 Determination of the source number N_s from inspection of (a) AIC and (b) MDL.

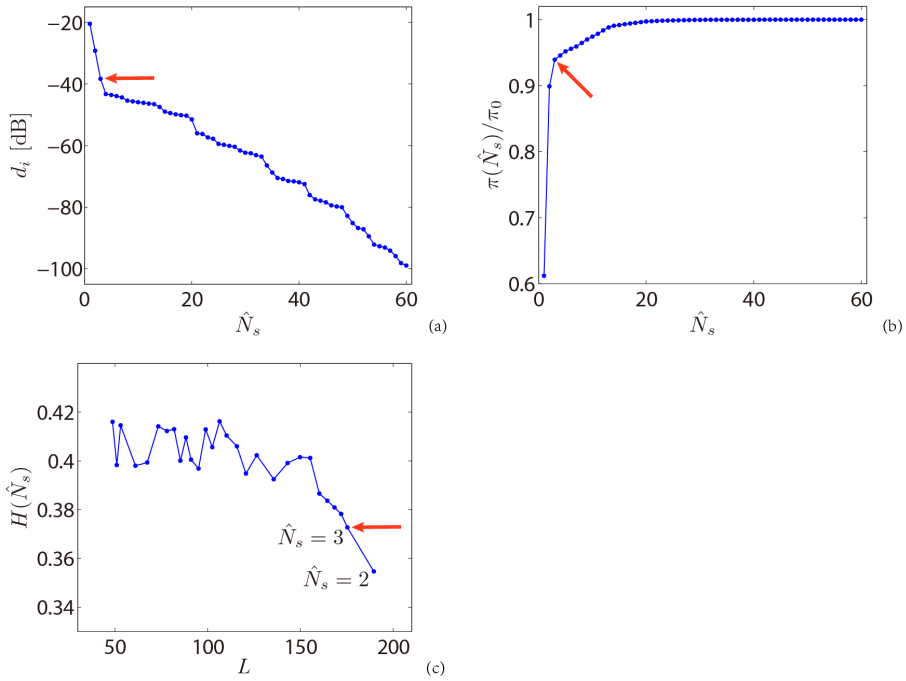

Fig. 6.60 Determination of the source number N_s from (a) the eigenvalue spectrum, (b) the cumulative power distribution, and (c) the entropic L-curve.

Among all proposed methods, CLSE is used here to separate the sources due to its better performance. The separation results and their comparison to actual sources are displayed in Figs. 6.61-6.62, respectively. Because it has the best radiation efficiency towards the array, the monopole was separated first, then the dipole, and finally the quadrupole. The separated sources have spatial shapes that closely match the actual ones in all cases; errors of reconstruction in terms of magnitude are all less than 3 dB. Indeed, the separated sources are again slightly more focused than the actual ones (because their compactness is forced), which explains a small overestimation of their magnitude in order to keep the same power. Note also that this numerical simulation demonstrates that sources can be separated even though they are totally overlapping in space.

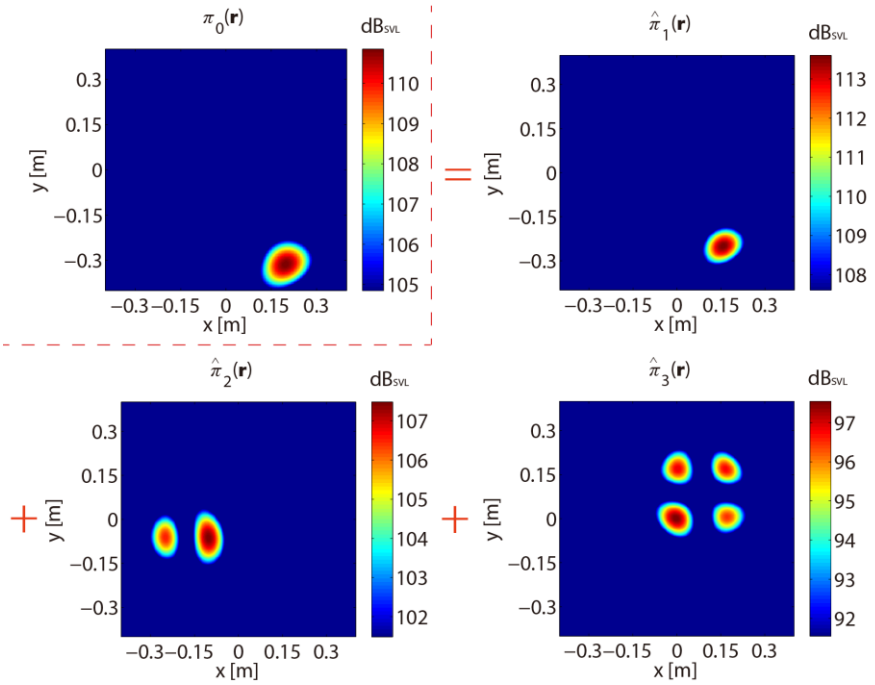

Fig. 6.61 Separated sources, $\hat{\pi}_i(\mathbf{r},\omega)$, i = 1, 2, 3, from the criterion of least spatial entropy.

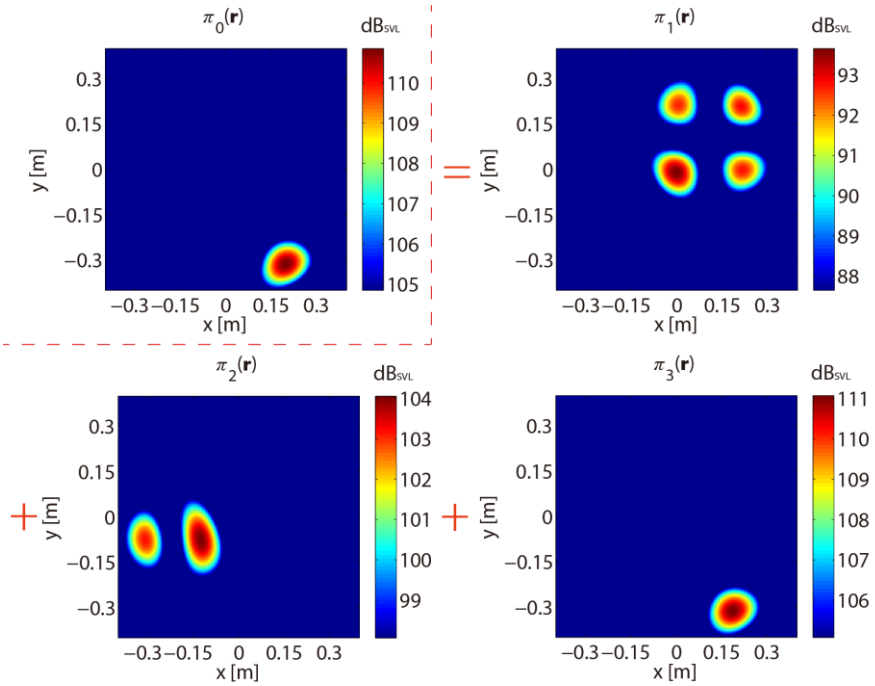

Fig. 6.62 Actual sources, $\pi_i(\mathbf{r},\omega)$, i = 1, 2, 3.

6.3 An industrial example

In addition to laboratory experiments, the proposed methods are illustrated here for the separation of sound sources produced by a Diesel engine, an industrial example which has for long attracted attention due to its importance – see e.g. [153]-[156] for the application of supervised separation to this problem. The research context and the full experimental apparatus is described in Ref. [8], which developed a methodology for separating the combustion, injection, and mechanical noises by making use of references and of the cyclostationary property of signals generated by a Diesel engine. Here, separation is achieved based on much weaker requirements, without references, (that is "blindly") and without the cyclostationary assumption (results are "stationarized" by averaging over time). The top side of the Diesel engine is the source surface of interest; the distance Z from the hologram plane is 18 cm. The array is a slice wheel with 84 microphones. The sampling frequency is 16.384 kHz. All results are presented at frequency 1250 Hz as considered in Ref. [8] in order to allow comparisons with the latter (the working frequency 1250 Hz is a good illustration where several sources radiate together, with similar levels, in the audible range).

6.3.1 Determination of the number of sources

As before, the number of active sources is first determined before trying to separate them. In practice, this may not be so obvious since a Diesel engine will radiate sound waves originating from a myriad of different sources, which are difficult to limit to a finite number. The information theoretic criteria seem to provide reasonable numbers of sources: $\hat{N}_s = 12$ for AIC and $\hat{N}_s = 11$ for MDL, as shown in Fig. 6.63. In such a scenario with 84 microphones and high spatial resolution – 2 mm – searching for an 11×11 or 12×12 matrix $\mathbf{V}$ would be numerically difficult. Furthermore, one may not expect a clear cut-off in the eigenvalue spectrum, nor a clear inflexion in the cumulative power distribution. This is clearly illustrated in Fig. 6.64(a). The source number may be set to 3, as suggested by the small gap between the third and the

fourth eigenvalues in Fig. 6.64(a). However, this would explain less than 85% of the overall power, as shown in Fig. 6.64(b). If at least 90% of the power had to be covered by the separated sources, then $\hat{N}_s = 7$ would probably be a better choice as the red arrow shows. The inflexion point in the entropic L-curve suggests the optimal source number around 8, which corresponds to 95.89% of the power, as shown in Fig. 6.64(c).

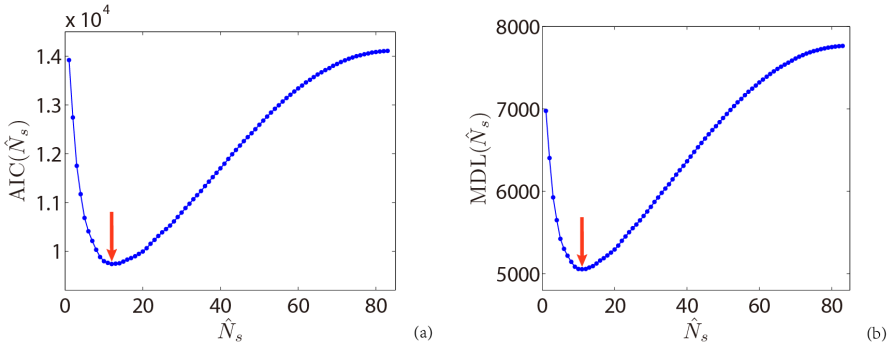

Fig. 6.63 Determination of the source number N_s from inspection of (a) AIC and (b) MDL (Diesel engine at 1250 Hz).

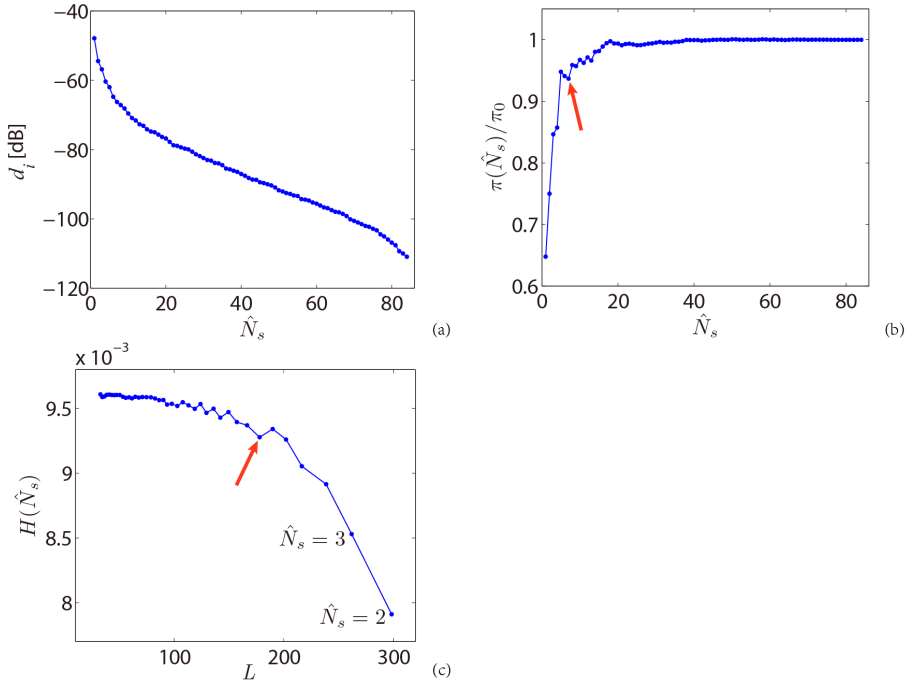

Fig. 6.64 Determination of the source number N_s from (a) the eigenvalue spectrum, (b) the cumulative power distribution, and (c) the entropic L-curve (Diesel engine at 1250 Hz).

6.3.2. Source separation from spatial criteria

Only the first four most physical separated sources returned by 2SD-JAD, 2SD-CG, CLSV, and CLSE with $\hat{N}_s = 8$ are displayed in Figs. 6.65, 6.67, 6.69 and 6.71, respectively. The spatial resolution is set to 2 mm for all the separated sources from the Diesel engine. Note that the indices of separated sources are determined in terms of the ranking of their power. It is seen that, by construction, joint statistical and spatial decorrelation (i.e. 2SD-JAD and 2SD-CG) returns less compact sources than the principle of least spatial complexity (i.e. CLSV and CLSE); as discussed hereafter, the latter are probably farther than the former from the actual (unknown) sources. The spatial distribution of the separated sources from CLSV and CLSE are almost the same, although there are some discrepancies in the amplitude. This evidences the presence of sources from different spatial origins.

According to experts (see also the discussion in Ref. [8]), the first and second sources correspond to combustion noise: for CLSE, the first source originates from the crank pulley and the second one involves part of the exhaust systems; for CLSV, the first two sources both involve the radiation of the crank pulley when excited by combustion and part of the exhaust systems; for 2SD-JAD and 2SD-CG, not only do the first two separated sources originate from the combustion noise, but also the third one, i.e. $\hat{\pi}_5(\mathbf{r})$ in Figs. 6.65 and 6.67, does. Interestingly, the four groups of results all show that combustion involves two "degrees of freedom" (i.e. uncorrelated modes). The fifth source from CLSE (corresponding to the fourth one from CLSV and the seventh ones from 2SD-JAD and 2SD-CG) is localized on the water pump and evidences radiation of some of its parts. The sixth source from CLSE (corresponding to the fifth one from CLSV) lays over the injection system of the Diesel engine and thus corresponds to injection noise. The other separated sources are not shown due to the difficulty of assigning them to definite physical origins. Unfortunately, the results

from 2SD (includes JAD and CG) lose the source originating from the injection noise, which again show some inferiority of this approach as compared to the criteria of least spatial complexity.

These results from CLSV and CLSE correlate well with the separation obtained in Ref. [8] with the help of references on the combustion and injection processes (first and sixth sources). They make possible the ranking of sources, an objective of fundamental importance to the engineer; by way of an example, Figs. 6.66, 6.68, 6.70, and 6.72 display all the partial powers of the separated sources, together with the residual power not explained by them, from the four methods.

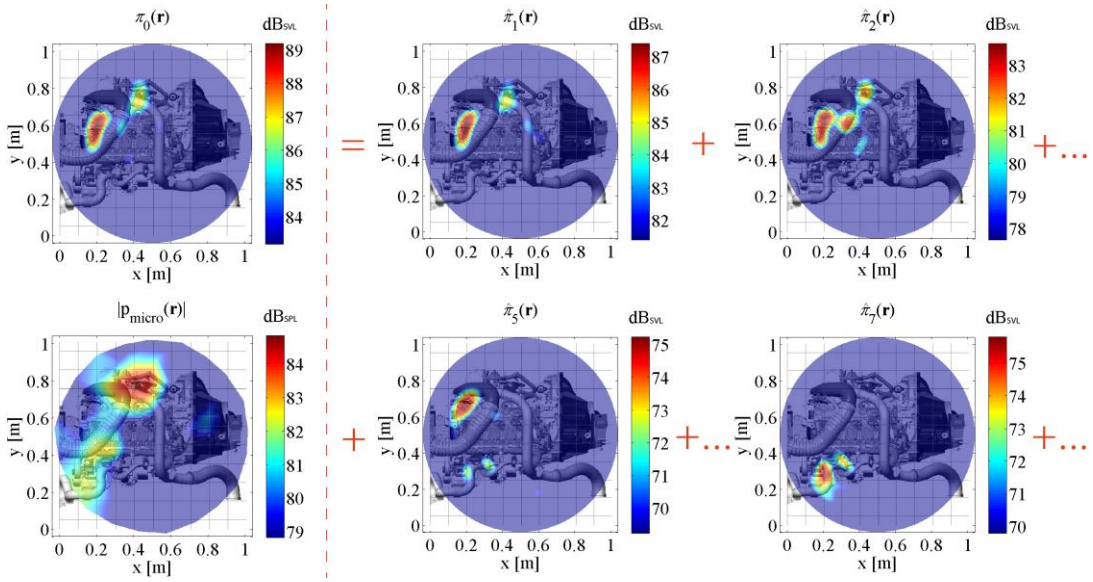

Fig. 6.65 Separated sources, $\hat{\pi}_i(\mathbf{r},\omega)$, i = 1,2,5,7, from joint statistical and spatial decorrelation with joint approximate diagonalization (Diesel engine at 1250 Hz).

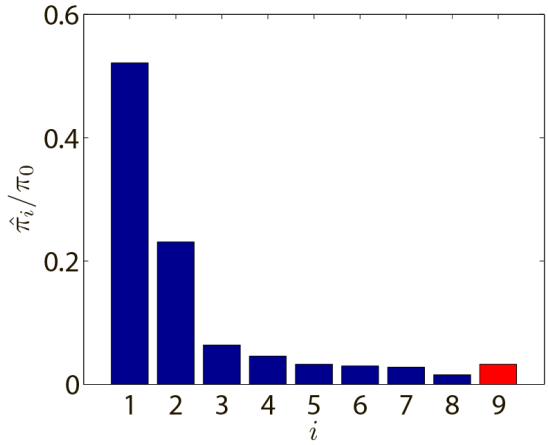

Fig. 6.66 Ratios of power of the separated sources and of the residual power (red bar) to the total power from joint statistical and spatial decorrelation with joint approximate diagonalization (Diesel engine at 1250 Hz).

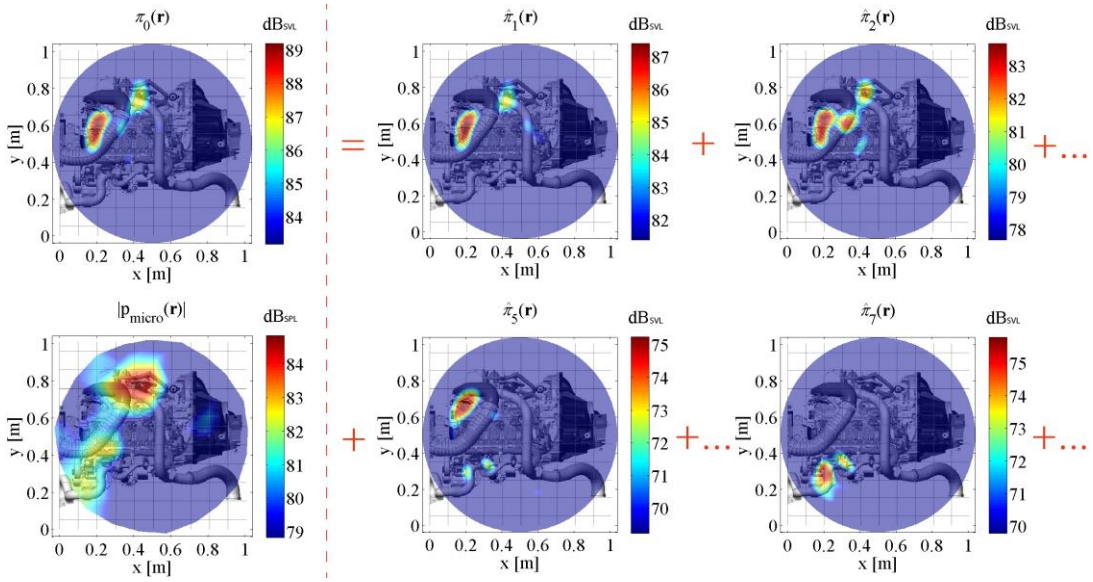

Fig. 6.67 Separated sources, $\hat{\pi}_i(\mathbf{r},\omega)$, i = 1,2,5,7, from joint statistical and spatial decorrelation with the conjugate gradient method (Diesel engine at 1250 Hz).

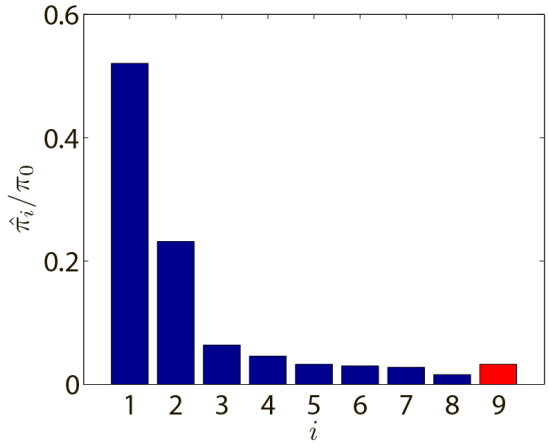

Fig. 6.68 Ratios of power of the separated sources and of the residual power (red bar) to the total power from joint statistical and spatial decorrelation with conjugate gradient method (Diesel engine at 1250 Hz).

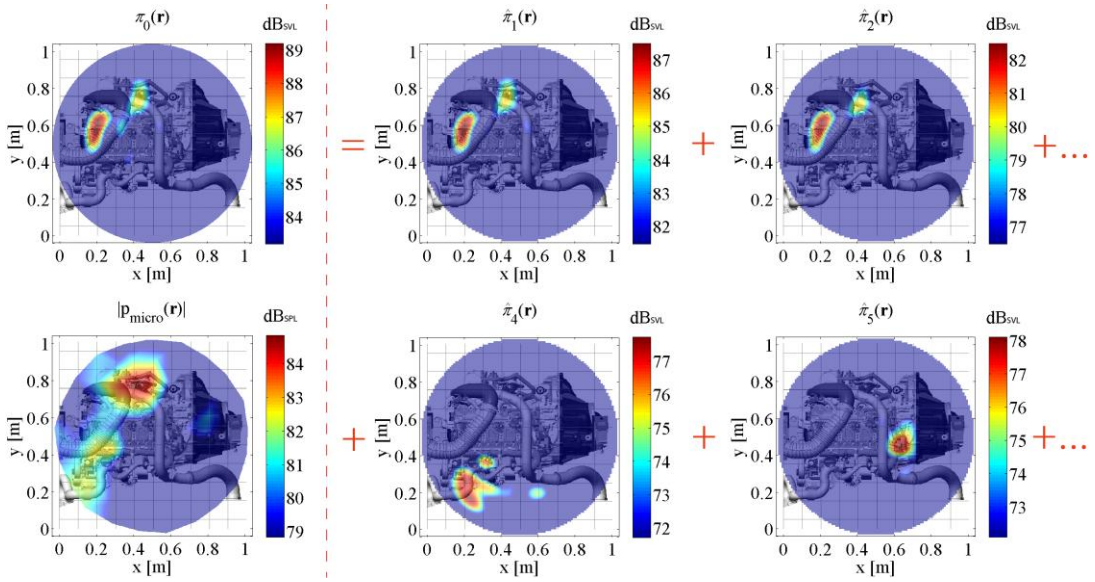

Fig. 6.69 Separated sources, $\hat{\pi}_i(\mathbf{r},\omega)$, i = 1,2,4,5, from the criterion of least spatial variance (Diesel engine at 1250 Hz).

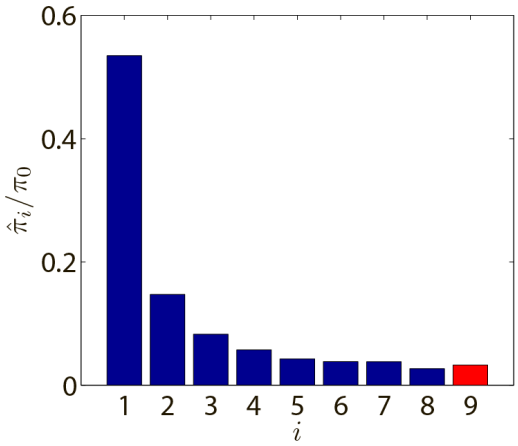

Fig. 6.70 Ratios of power of the separated sources and of the residual power (red bar) to the total power from the criteria of least spatial variance (Diesel engine at 1250 Hz).

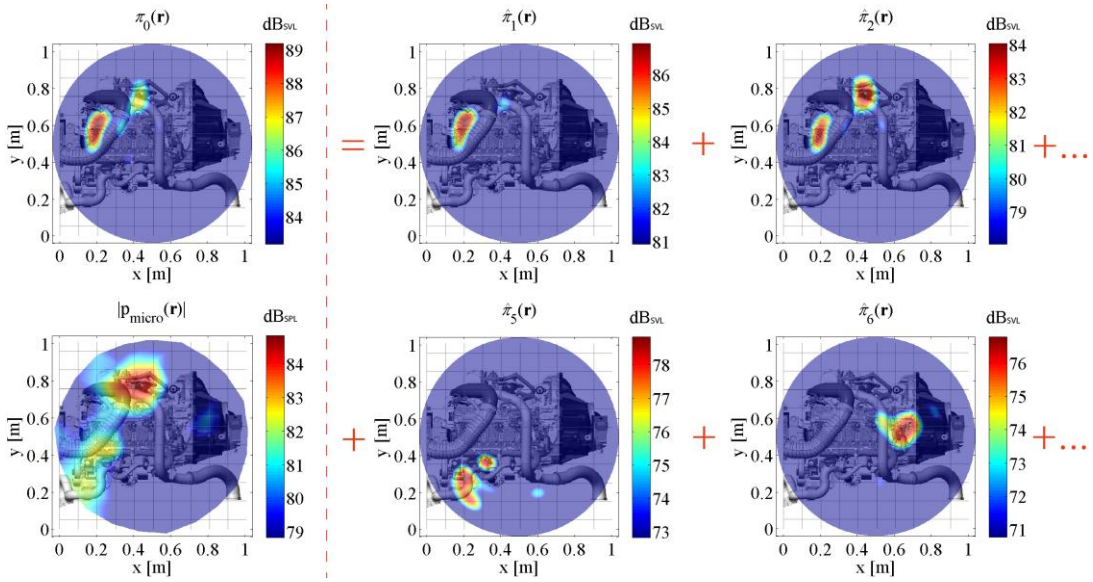

Fig. 6.71 Separated sources, $\hat{\pi}_i(\mathbf{r},\omega)$, i = 1,2,5,6, from the criterion of least spatial entropy (Diesel engine at 1250 Hz).

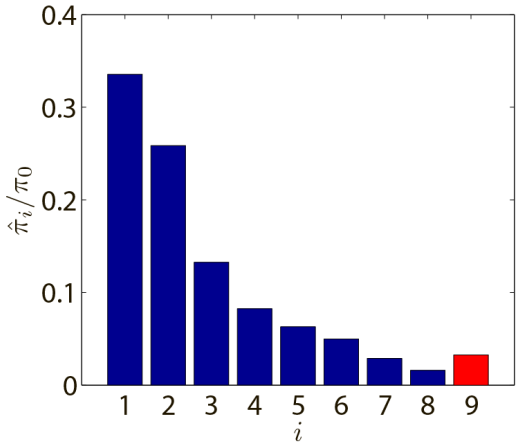

Fig. 6.72 Ratios of power of the separated sources and of the residual power (red bar) to the total power from the criteria of least spatial entropy (Diesel engine at 1250 Hz).

6.3.3 Source separation from ranking eigen elements according to increasing spatial entropy

The cluster of virtual sources of interest is limited to the first 25 ones, after contrasting the three curves in Fig. 6.64. The spatial entropies of the first 25 virtual sources are depicted in Fig. 6.73. According to increasing spatial entropies, the virtual sources numbered as 1, 2, 3, 5, 14, 21, 4 and 16 are chosen to take part in the following optimization operation. After ranking the elements in matrices **D** and **U** according to the indices of the selected virtual sources and then applying CLSE, the first seven most physical sources shown in Fig. 6.74 were obtained. Remarkably, not only are the first four sources the same as in Fig. 6.71, but also the other three smaller sources, which overlap the first four sources in space, are located. The fifth source, i.e. $\hat{\pi}_6(\mathbf{r})$, also comes from the injection noise, but they are radiated by different mechanical components. $\hat{\pi}_7(\mathbf{r})$ is the contribution from the combustion noise, but from a different component, which might be the fake copy of $\hat{\pi}_1(\mathbf{r})$. The last one – i.e. $\hat{\pi}_8(\mathbf{r})$, involves another part of the exhaust systems. This example demonstrates the availability of CLSE to separate overlapping sources even in difficult industrial

context. Figure 6.75 displays all partial power of the separated sources plus the residual power. In this improved scenario, 95.07% of the overall power is involved, which is a little lower than the former one – 96.74% for CLSV in Fig. 6.70 and for CLSE in Fig. 6.72.

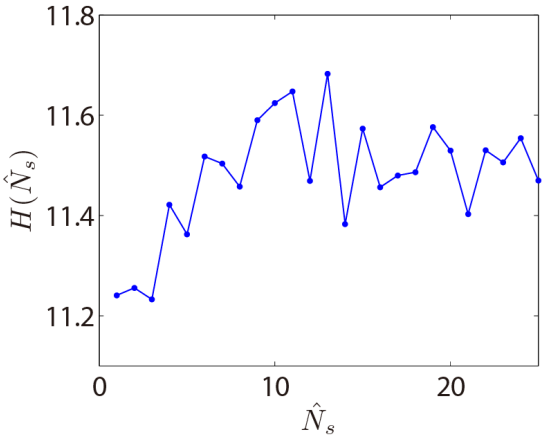

Fig. 6.73 Spatial entropy of the first 25 virtual sources (Diesel engine at 1250 Hz).

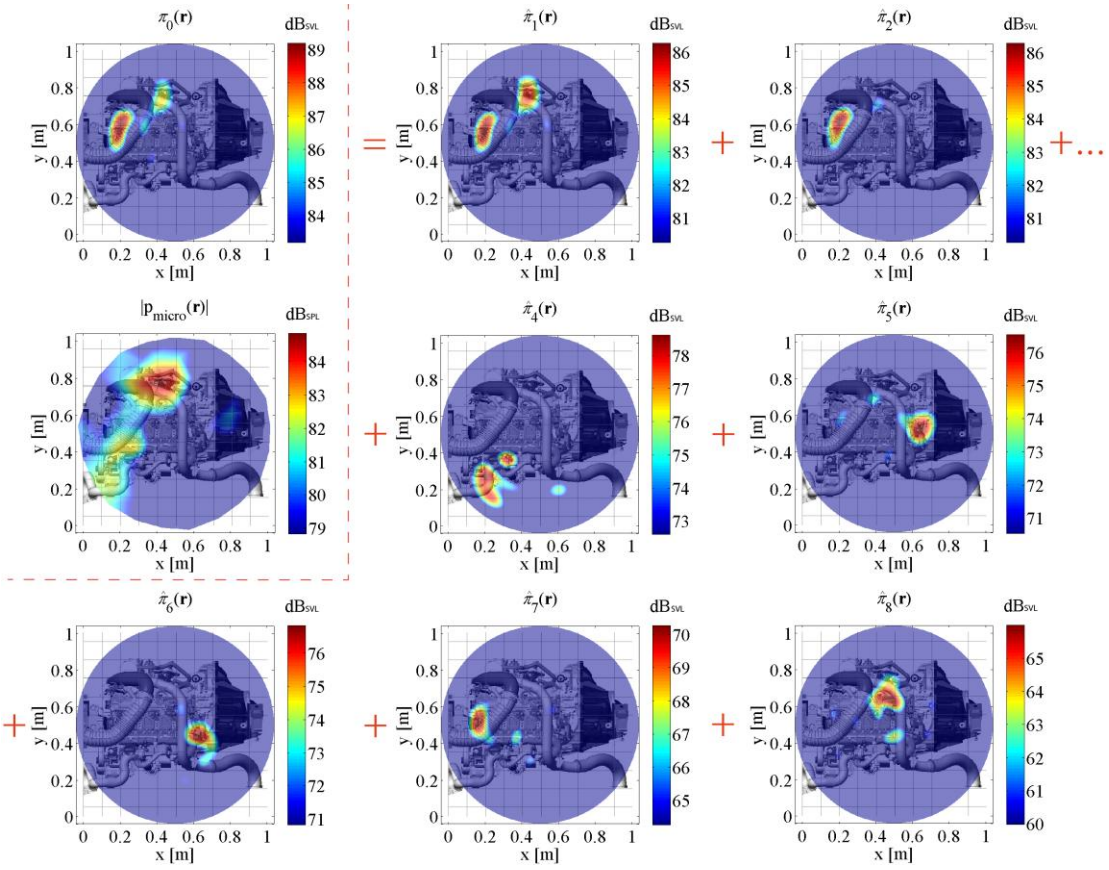

Fig. 6.74 Separated sources, $\hat{\pi}_i(\mathbf{r},\omega)$, i = 1,2,4,5,6,7,8, from the criterion of least spatial

entropy, ranked according to increasing spatial entropy (Diesel engine at 1250 Hz).

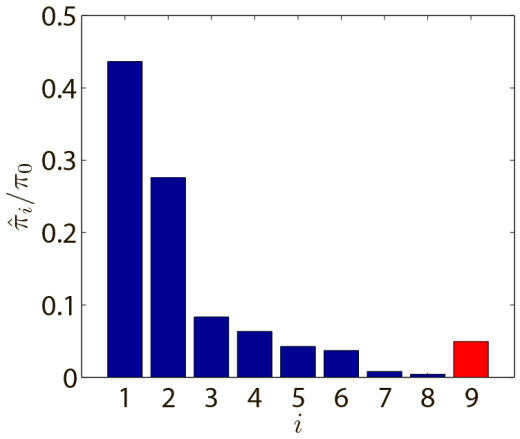

Fig. 6.75 Ratios of power of the separated sources (ranked according to their spatial entropies) and of the residual power (red bar) to the total power (Diesel engine at 1250 Hz).

Based on all separated results, the criterion of least spatial entropy is recommended hereafter. To separate very small sound sources, the order of the eigen elements in matrix **D** should be rearranged according to increasing spatial entropy (or spatial variance) of virtual sources first.

Chapter 7 Parametric analysis

The present chapter is dedicated to analyze how the various experimental parameters affect BSS. Based on the results obtained in Chapter 6, it is clear that CLSE owns the best performance for blind separation of sound sources. Therefore, in the present chapter, CLSE is used systematically despite its computational cost as compared to the other methods.

7.1 Robustness to the number of sources

It is shown herein that the proposed algorithm – CLSE is robust to the estimated number of sources. This is demonstrated by means of three indicators: namely, the power, the entropy, and the spatial localization of the most intense separated sources are all shown to be little affected by how many other sources are considered in the mixture. The robustness to the number of sources is double-checked by laboratory experiments and on the industrial example involving sound sources produced by the Diesel engine.

7.1.1 Laboratory experiments in the near-field

By means of an example, Fig. 7.1 displays the first separated source from CLSE, in the near-field, when the number $\hat{N}_s$ varies from 2 to 6, where $\pi_1(\mathbf{r})$ stands for the actual first source distribution and $\hat{\pi}_1^{[\hat{N}_s]}(\mathbf{r})$ for its estimation assuming $\hat{N}_s$ sources. The actual source number is 4. It is seen that the estimated sources look the same, both in magnitude and in shape, when increasing the value of $\hat{N}_s$.

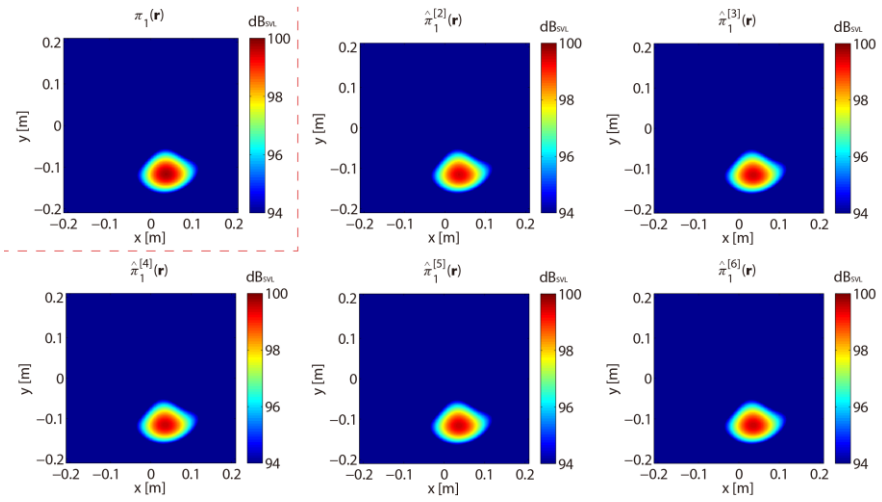

Fig. 7.1 The first separated source, $\hat{\pi}_1^{[\hat{N}_s]}(\mathbf{r})$, versus the estimated source number, $\hat{N}_s$=1,…,6 (criterion of least spatial entropy in the near-field at 833 Hz).

The ratio of powers and the ratio of spatial entropies of the first separated source to the actual one are illustrated in Figs. 7.2(a) and 7.2(b), respectively, confirming an excellent stability with respect to the estimated source number. Finally, the localization of the highest peak in the first separated source versus the estimated source number is depicted in Fig. 7.3. The error in localization is seen to be 0.001m, which corresponds to the spatial resolution of the reconstructed sources.

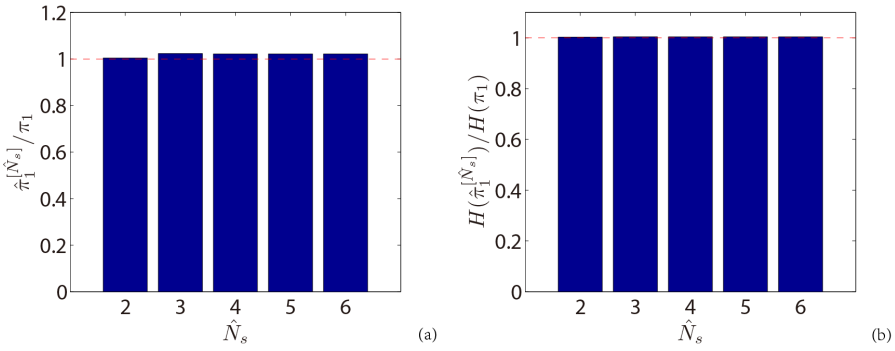

Fig. 7.2 (a) Power ratio of the first separated source to its actual value versus the estimated source number $\hat{N}_s$. (b) Ratio of spatial entropy of the first separated source to its actual value versus the estimated source number $\hat{N}_s$ (near-field at 833 Hz).

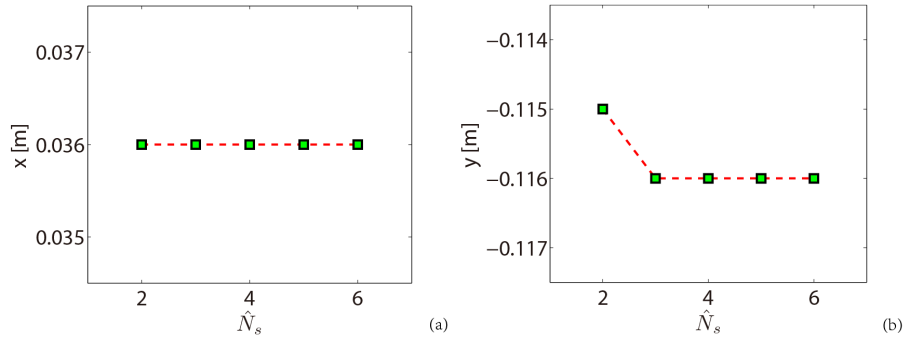

Fig. 7.3 Positions of the highest peak in the first separated source versus the estimated source number $\hat{N}_s$: (a) position on the x-axis and (b) position on the y-axis (near-field at 833 Hz).

7.1.2 Laboratory experiments in the far-field

The first separated source from CLSE is depicted in Fig. 7.4, while the source number $\hat{N}_s$ varies from 2 to 6 in the far-field at 2437 Hz. The actual number of sources is 4. All the sources appear very similar, if not identical, both in amplitude and in shape.

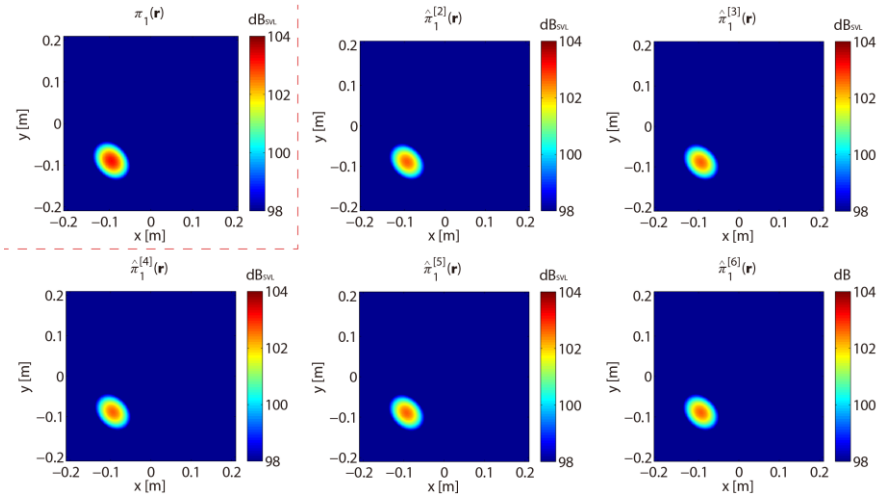

Fig. 7.4 The first separated source, $\hat{\pi}_1^{[\hat{N}_s]}(\mathbf{r})$, versus the estimated source number $\hat{N}_s$=1,…,6 (criterion of least spatial entropy in the far-field at 2437 Hz).

Similarly to Fig. 7.2, the ratio of powers and the ratio of spatial entropies of the first separated source to the actual one in the far-field are illustrated in Figs. 7.4(a) and

7.4(b), respectively, confirming again an excellent stability to the estimated source number. The powers of all separated sources are lower than the reference, but they have identical values. The difference in the powers between the separated sources and the actual source is likely to come from a limitation of the backpropagation process in the far-field. Afterwards, the localization of the highest peak in the sources versus the estimated source number is depicted in Fig. 7.6. Remarkably, there is no bias in positions of x- and y-axes.

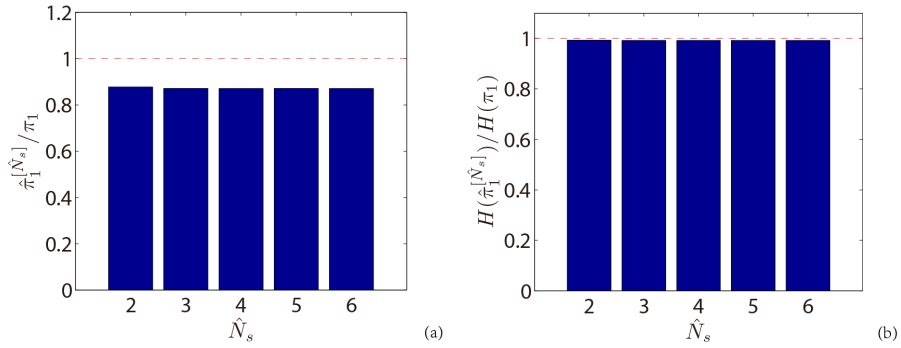

Fig. 7.5 (a) Power ratio of the first separated source to its actual value versus the estimated source number $\hat{N}_s$. (b) Ratio of spatial entropy of the first separated source to its actual value versus the estimated source number $\hat{N}_s$ (far-field at 2437 Hz).

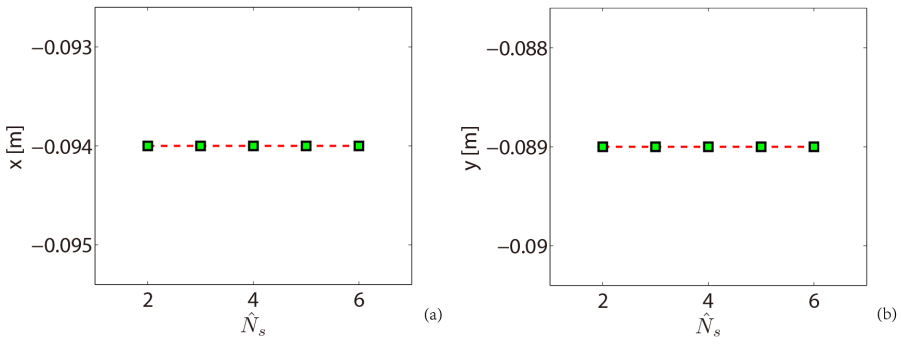

Fig. 7.6 Positions of the highest peak in the first separated source versus the estimated source number $\hat{N}_s$: (a) position on the x-axis and (b) position on the y-axis (far-field at 2437 Hz).

7.1.3 Industrial example

To verify the robustness to the number of sources in an industrial application, CLSE is then applied to blindly separate sound sources radiated from the Diesel engine at 1250 Hz with $\hat{N}_s$ varying from 4 to 10 by steps of 2. Here, two sources of interest, i.e. the second and third ones (i.e. $\hat{\pi}_2(\mathbf{r})$ and $\hat{\pi}_5(\mathbf{r})$) in Fig. 6.71, are picked out, and then coined as $\pi_1(\mathbf{r})$ and $\pi_2(\mathbf{r})$ hereafter. The separated results from CLSE are illustrated for $\pi_1(\mathbf{r})$ in Fig. 7.7 and for $\pi_2(\mathbf{r})$ in Fig. 7.8.

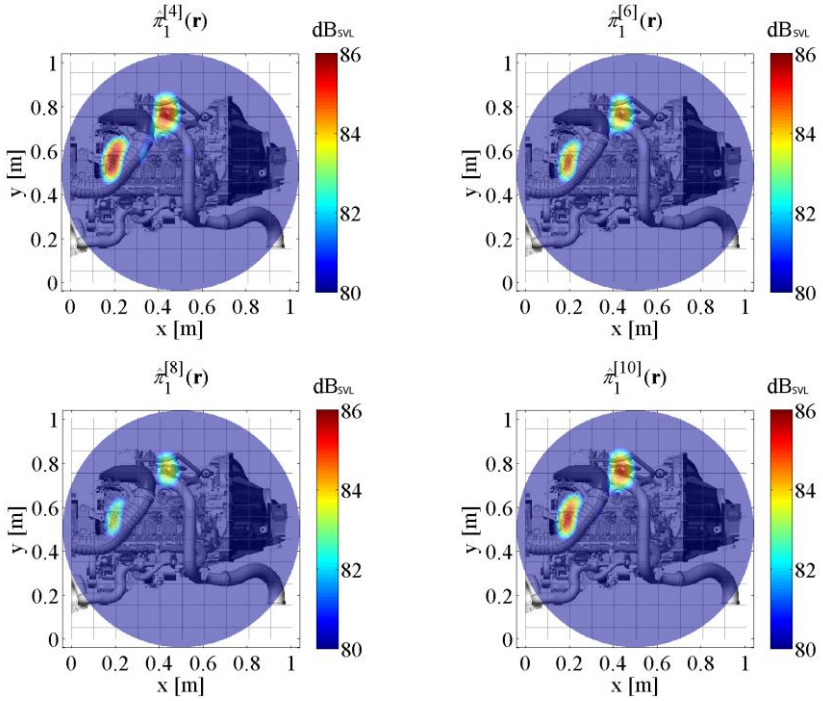

Fig. 7.7 The first separated source, $\hat{\pi}_1^{[\hat{N}_s]}(\mathbf{r})$, versus the estimated source number $\hat{N}_s$ = 4,6,8,10 (criterion of least spatial entropy – Diesel engine at 1250 Hz).

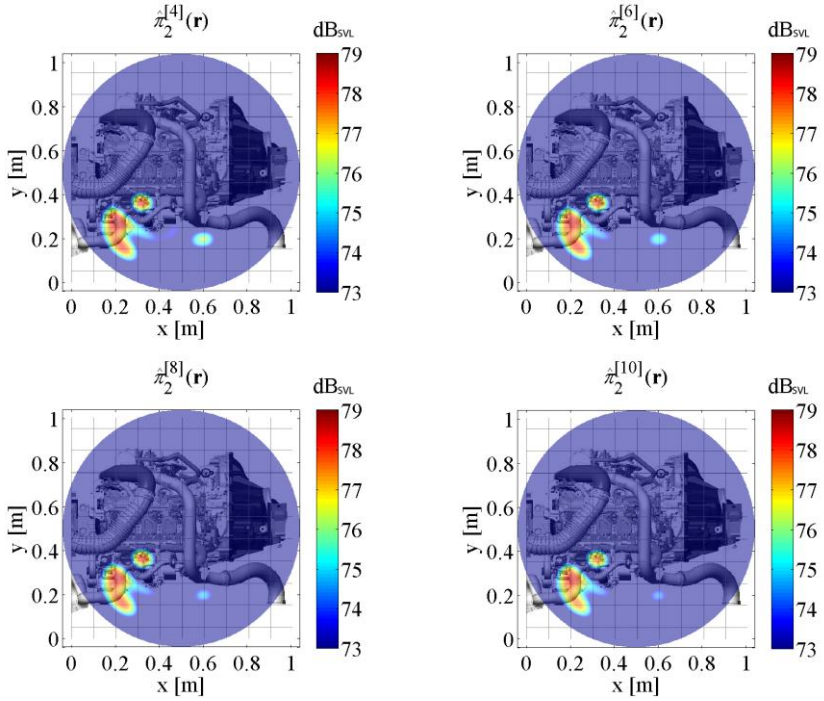

Fig. 7.8 The second separated source, $\hat{\pi}_2^{[\hat{N}_s]}(\mathbf{r})$, versus the estimated source number $\hat{N}_s$=4,6,8,10 (criterion of least spatial entropy – Diesel engine at 1250 Hz).

Similarly, Fig. 7.9 displays the ratio of powers and the ratio of spatial entropies of the first separated sources to their mean value and Fig. 7.10 for the second source. No matter the first separated source or the second one, the spatial entropy nearly keeps constant, although their powers fluctuate with the value of $\hat{N}_s$. That is to say the shapes of the separated sources are robust to the number of sources (as shown in Figs. 7.7-7.8), which is useful to identify the sources and locate their positions. The power of the first separated source varies more than the second one because the first source overlaps with other 7 sources in space (see Fig. 6.71). However, the bias among the first separated sources is less than 2dB in the amplitude of its highest peak, which demonstrates CLSE robustness to the number of sources from the energy point of view.

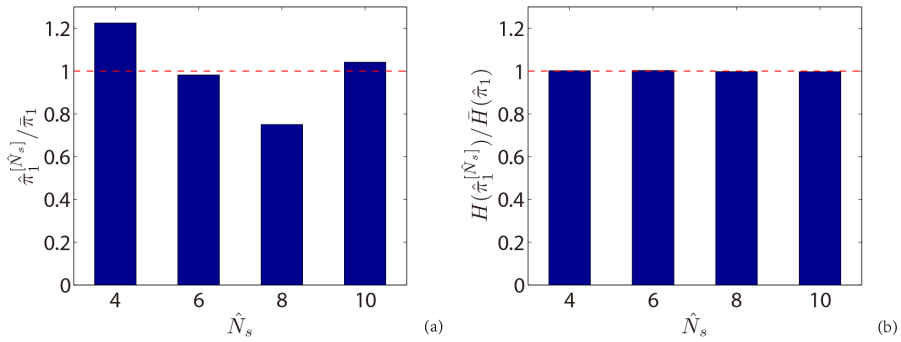

Fig. 7.9 (a) Power ratio of the first separated source to its mean value versus the estimated source number $\hat{N}_s$. (b) Ratio of spatial entropy of the first separated source to its mean value versus the estimated source number $\hat{N}_s$ (Diesel engine at 1250 Hz). (The bar '¯' denotes the average operation on energy for Fig. 7.9(a) and on space for Fig. 7.9(b))

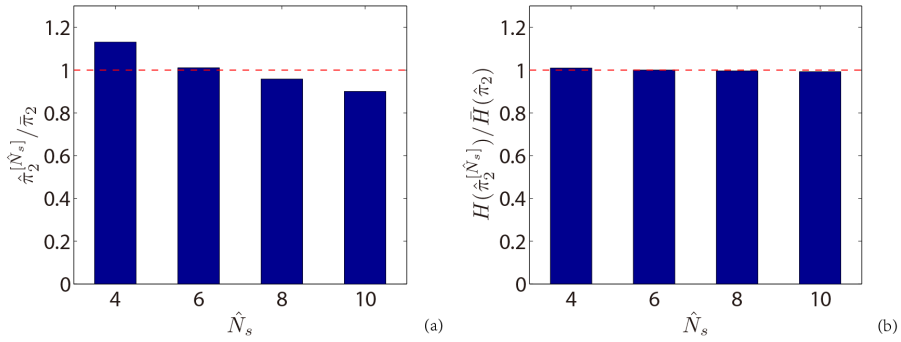

Fig. 7.10 (a) Power ratio of the second separated source to its mean value versus the estimated source number $\hat{N}_s$. (b) Ratio of spatial entropy of the second separated source to its mean value versus the estimated source number $\hat{N}_s$ (Diesel engine at 1250 Hz).

The localization of the highest peak versus the estimated source number is depicted for the first separated source in Fig. 7.11 and for the second one in Fig. 7.12. The highest peak of $\hat{\pi}_1^{[4]}(\mathbf{r})$ in amplitude, locating in the left hand side of panels in Fig. 7.7, comes from the combustion noise, and the other peak corresponds to part of the exhaust system. The highest peak of $\hat{\pi}_2^{[4]}(\mathbf{r})$ in amplitude occupies much larger place than the next highest one as shown in Fig. 7.8. According to the aforementioned content, $\pi_2(\mathbf{r})$ originates from the water pump and its first peak mainly concentrates

on the tube of the water pump. Inspecting Figs. 7.11-7.12, there are very tiny errors on the localizations of the highest peaks of the two separated sources. Moreover, the maximum error on the localization is 16 mm on the direction of y-axis as presented in Fig. 7.11(b). The robustness of CLSE to the estimated number of sources is once again confirmed by the localizations of the two sources – $\pi_1(\mathbf{r})$ and $\pi_2(\mathbf{r})$.

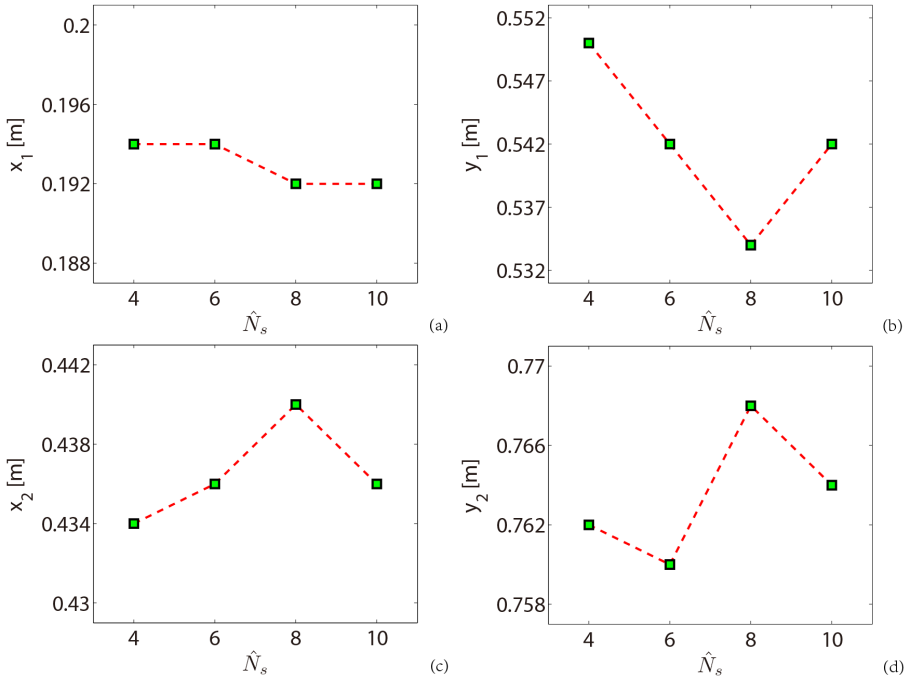

Fig. 7.11 Positions of the highest peak in the first separated source versus the estimated source number $\hat{N}_s$: (a) position of the highest peak on the x-axis, (b) position of the highest peak on the y-axis, (c) position of the next highest peak on the x-axis, and (d) position of the next highest peak on the y-axis (Diesel engine at 1250 Hz).

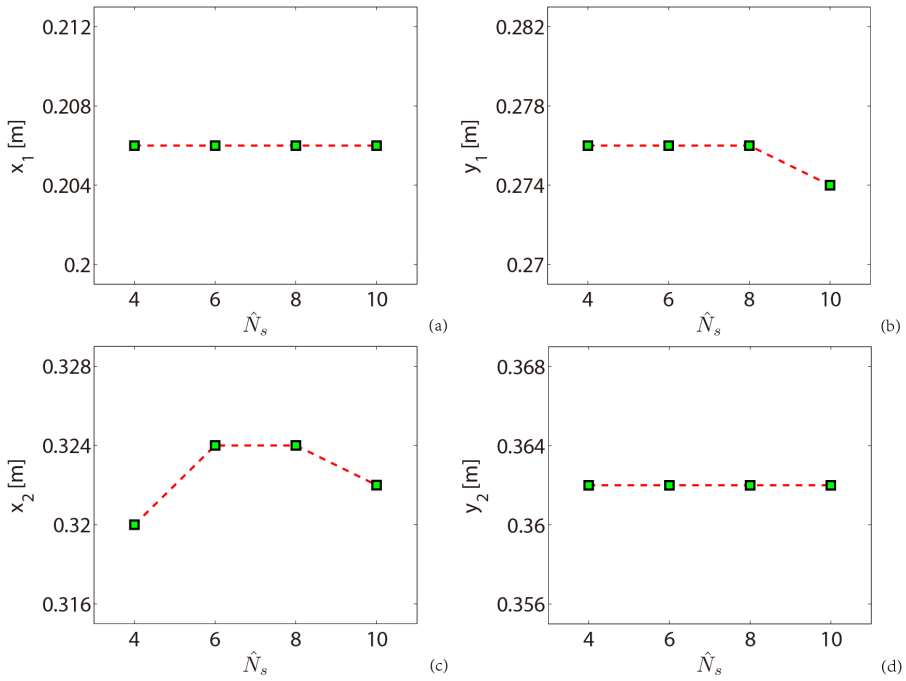

Fig. 7.12 Positions of the highest peak in the second separated source versus the estimated source number $\hat{N}_s$: (a) position of the highest peak on the x-axis, (b) position of the highest peak on the y-axis, (c) position of the next highest peak on the x-axis, and (d) position of the next highest peak on the y-axis (Diesel engine at 1250 Hz).

7.2 Optimal distance for backpropagation

According to Huygens principle and to the mechanism of backpropagation, there is an actual distance Z_0 between the reconstruction plane and the array of microphones, where the separated sources are the most compact. This actual distance is defined as the minimum of the spatial entropy (or the spatial variance) of the full reconstructed source field as a function of the range distance. Obviously, such an optimum should be close to the nominal distance between the source plane and the array. In the present section, BSS at the actual distance is investigated by means of the laboratory experiments and the industrial example.

7.2.1 Laboratory experiments in the near-field

The spatial entropy of the full reconstructed source field versus the distance Z is

displayed in Fig. 7.13. The minimum of the curve corresponds to Z_{opt} = 11.2 cm, which is close to the nominal distance Z_0 = 10 cm, and coincides with the conclusion on the selection of reconstruction plane proposed by J. Hald [13]. Therefore, there is definitely a minimum corresponding to the most compact spatial distribution. Figure 7.14 displays all separated sources on the actual source plane at Z_{opt} = 11.2 cm. For comparison, separation results are compared at smaller and larger range distances in Fig. 7.15 and Fig. 7.16, respectively. Despite severe distortion in shape, CLSE can still successfully separate the sources in the case of large underestimation (Z = 5cm) or overestimation (Z = 22cm) of the nominal distance. This means CLSE is robust to the range distance Z in the near-field, even though the reconstructed sources suffer serious distortion and bias.

Furthermore, the coordinates of the hottest points of the four separated sources versus the distance Z are listed in Tab. 7.1. Among all distances Z, the separated results on the actual plane Z_{opt} provides the most accurate localization of the sources (see Tab. 7.1). The maximum position errors are respectively 2.3 cm in the x-axis and 3 cm in the y-axis with respect to the localizations of the sources on the actual plane, which are both lower than half the spacing D_0 = 12cm between the sources and one order of magnitude smaller than the wave length λ = 41.2cm at 833 Hz.

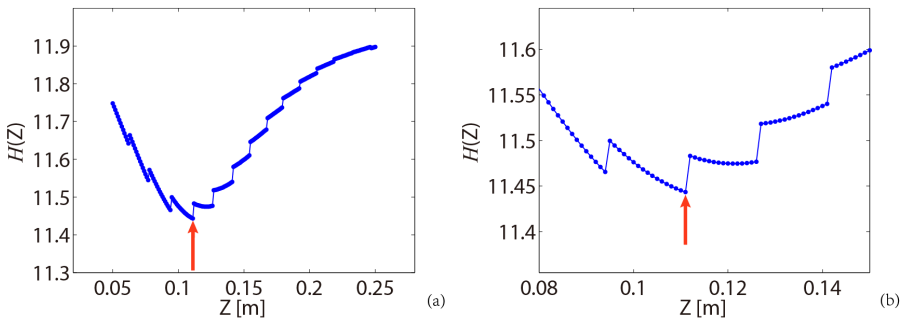

Fig. 7.13 (a) Spatial entropy of the full reconstructed source field versus range distance Z (near-field at 833 Hz). (b) Zoom around the actual distance.

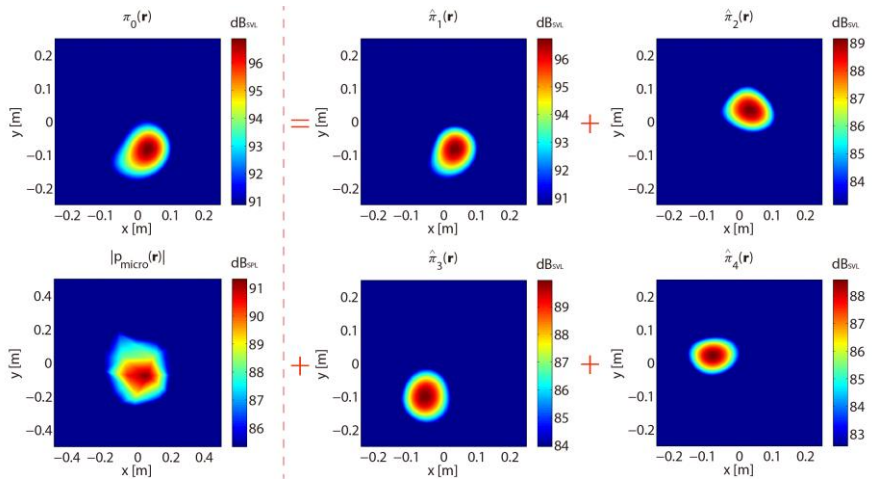

Fig. 7.14 Separated sources, $\hat{\pi}_i(\mathbf{r},\omega)$, $i = 1,\ldots,4$, from the criterion of least spatial entropy, at the actual range distance (near-field at 833 Hz).

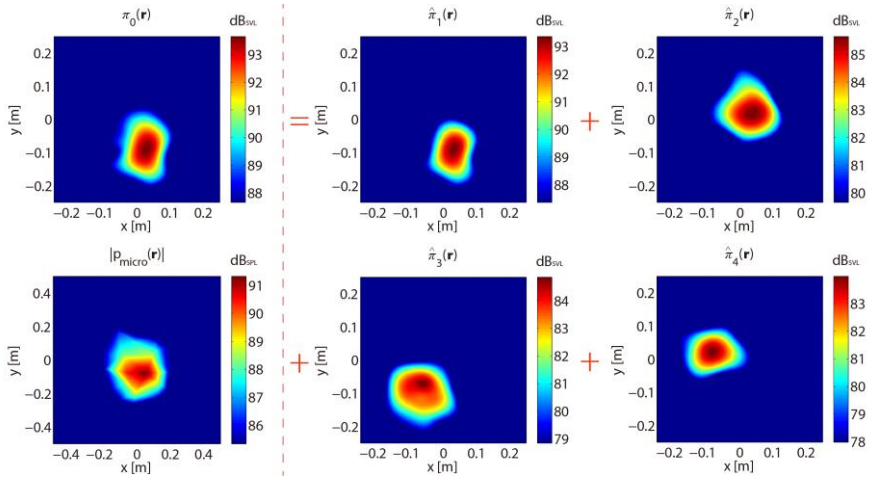

Fig. 7.15 Separated sources, $\hat{\pi}_i(\mathbf{r},\omega)$, $i = 1,\ldots,4$, from the criterion of least spatial entropy, at range distance $Z = 5$ cm (near-field at 833 Hz).

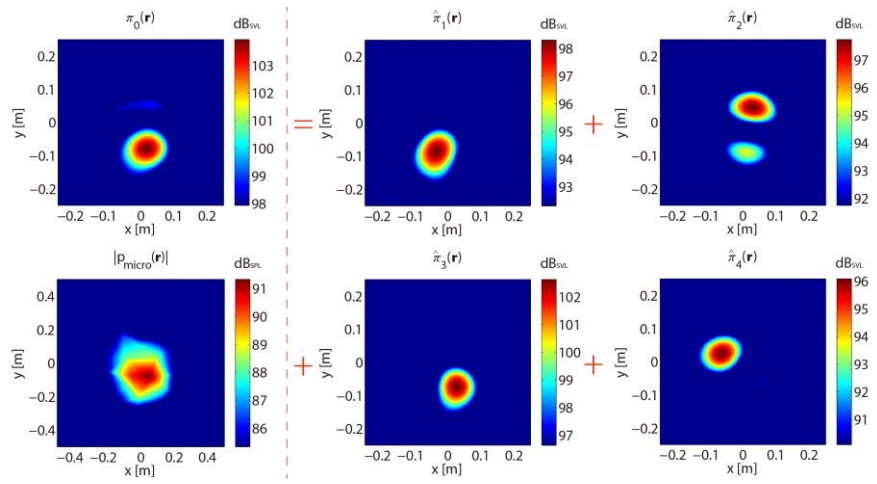

Fig. 7.16 Separated sources, $\hat{\pi}_i(\mathbf{r},\omega)$, $i = 1,\ldots,4$, from the criterion of least spatial entropy, at range distance $Z = 22$ cm (near-field at 833 Hz).

	s_1	s_2	s_3	s_4
$Z = 5$ cm	(-8.3, 2.3)	(3.8, 1.8)	(-6, -7)	(3.2, -8.8)
$Z = 10$ cm	(-8.3, 2.3)	(3.3, 3.4)	(-5.7, -10.2)	(3.5, -8.2)
$Z = 11.2$ cm	(-8.1, 2.3)	(3.2, 3.5)	(-5.6, -10)	(3.4, -8.2)
$Z = 22$ cm	(-6.4, 2.5)	(3, 4.5)	(-3.3, -8.5)	(2.7, -7.4)

Tab. 7.1 Coordinates (x, y) (in centimeter) of the highest peaks of four separated sources versus different range distances Z (near-field at 833 Hz).

7.2.2 Laboratory experiments in the far-field

In the far-field, the curve connecting the spatial entropy to the range distance Z is displayed in Fig. 7.17. Here, the actual distance is identical to the nominal one, i.e $Z_{opt} = Z_0 = 100$ cm. The corresponding separation results were displayed in Fig. 6.42 (with the radius of aperture function, $R = 21$ cm). Similarly, the separated sources for smaller and larger range distances are depicted in Figs. 7.18 and 7.19, respectively. In Fig. 7.19, the lowest peak of the fourth source relates to the first actual source, yet the source distribution seems dominated by the same hot point as in the third separated source. The two figures demonstrate that CLSE is robust to the distance Z from the

far-field point of view. Table 7.2 records the coordinates of the hottest points of the four separated sources with respect to different values of the range distance Z. The maximum bias on position is 3.7 cm on the x-axis and 3.2 cm on the y-axis, which is also lower than the half of the nominal spacing D_0 = 18cm in the far-field.

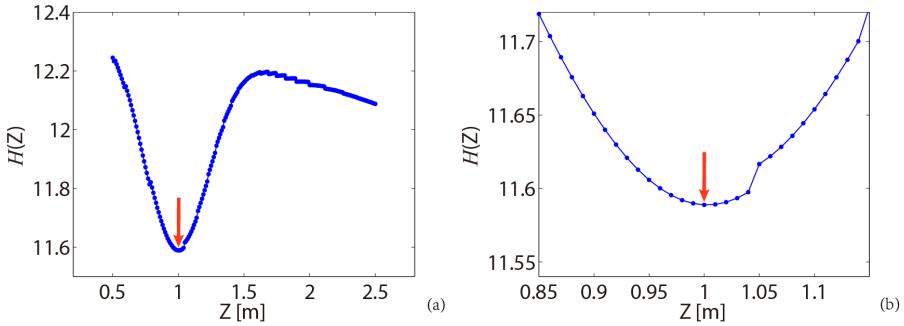

Fig. 7.17 (a) Spatial entropy of the full reconstructed source field versus the range distance Z (far-field at 2437 Hz). (b) Zoom around the actual distance.

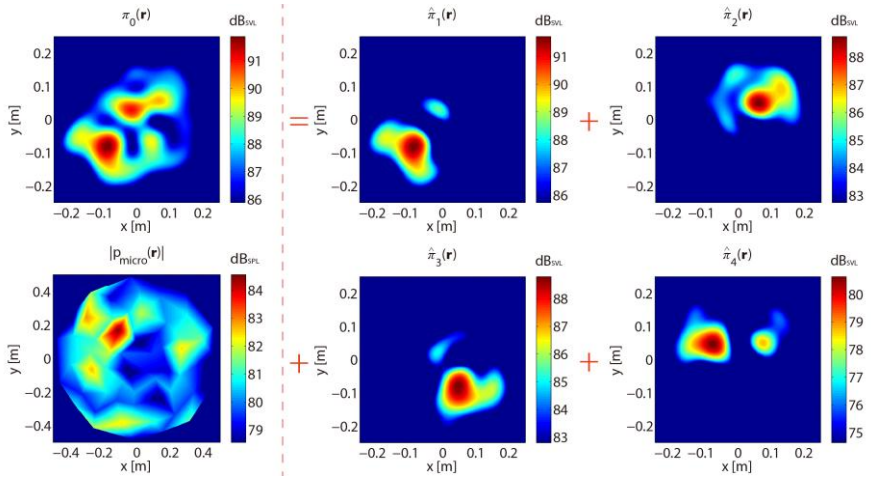

Fig. 7.18 Separated sources, $\hat{\pi}_i(\mathbf{r},\omega)$, i = 1,…,4, from the criterion of least spatial entropy, at range distance Z = 60 cm (far-field at 2437 Hz).

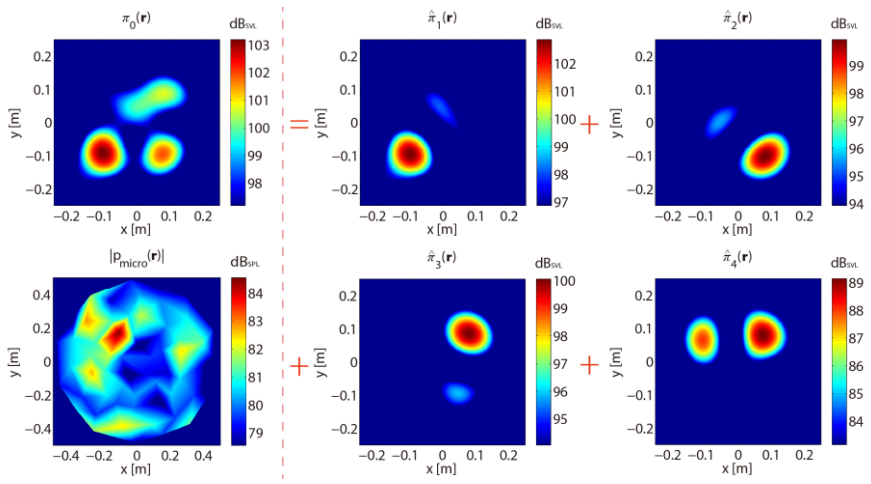

Fig. 7.19 Separated sources, $\hat{\pi}_i(\mathbf{r},\omega)$, $i = 1,\ldots,4$, from the criterion of least spatial entropy, at range distance $Z = 180$ cm (far-field at 2437 Hz).

	s_1	s_2	s_3	s_4
$Z = 60$ cm	(-7.5, 4.6)	(6, 5)	(-8.6, -8.1)	(4.8, -8.2)
$Z = 100$ cm	(-11.2, 7.8)	(6.8, 7)	(-10.5, -9.9)	(7, -9.7)
$Z = 180$ cm	(-10.6, 6.4)	(8, 8.6)	(-9.9, -9.5)	(8.2, -10.2)

Tab. 7.2 Coordinates (x, y) (in centimeter) of the highest peaks of four separated sources versus different range distances Z (far-field at 2437 Hz).

7.2.3 An industrial example

The robustness to the range distance Z between the reconstruction plane and the array of microphones is now verified on the Diesel engine example. In this case, the actual distance is found as $Z_{opt} = 21.3$ cm, as shown in Fig. 7.20, which is different slightly from the "nominal" distance $Z_0 = 18$ cm since the latter is known with a limited precision and actually represents an average distance to a non planar source domain . Figure 7.21 displays the primary four sources radiated from the Diesel engine. Contrasting them with those in Fig. 6.71, the separated sources on the actual plane appear more compact, especially the first and sixth ones. This is then compared with an underestimated range distance – $Z = 9$ cm (see Fig. 7.22) – and an

overestimated distance – Z = 27 cm (see Fig. 7.23). With varying the distance Z, the shape of sources zooms in and out, but overall CLSE can provide stable results. The coordinates of the hottest points of the four separated sources are given in Tab. 7.3. The maximum errors in localization is 6.6 cm on the x-axis and 4 cm on the y-axis relatively to the corresponding sources at the actual distance.

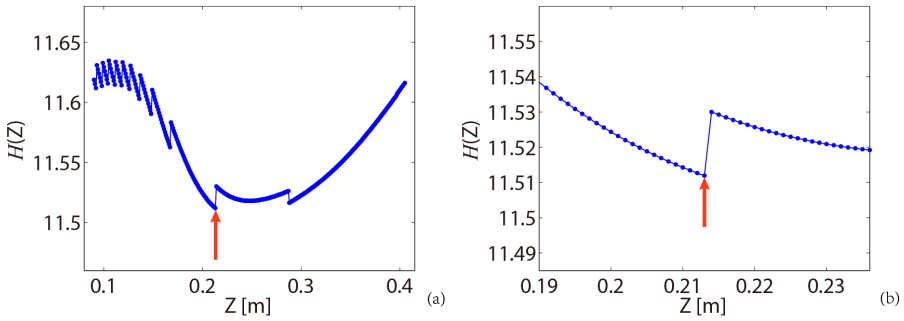

Fig. 7.20 (a) Spatial entropy of the full reconstructed source field versus the range distance Z (Diesel engine at 1250 Hz). (b) Zoom around the actual distance.

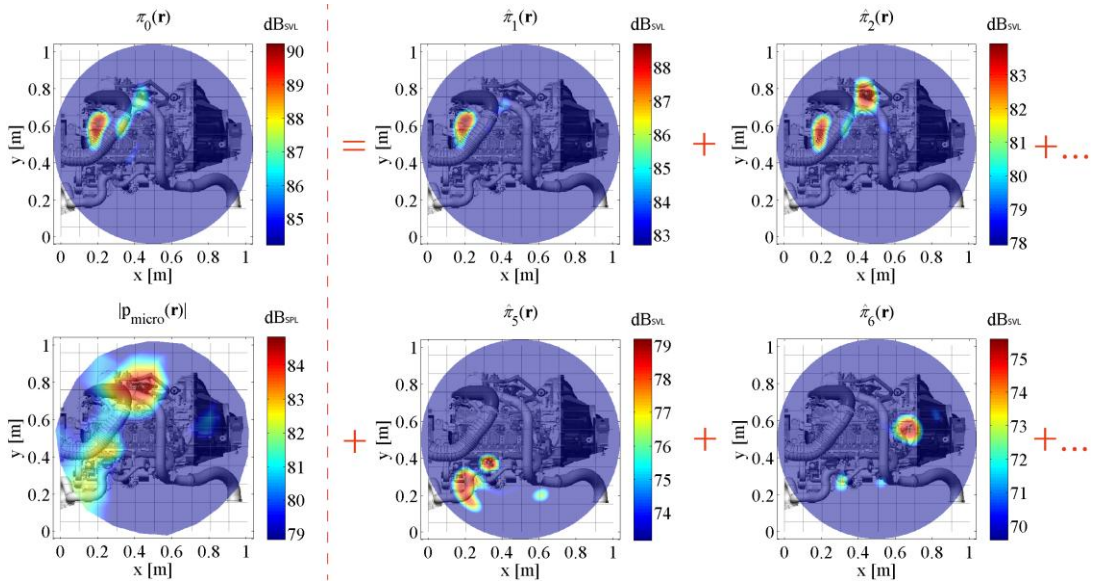

Fig. 7.21 Separated sources, $\hat{\pi}_i(\mathbf{r},\omega)$, i = 1,…,8, from the criterion of least spatial entropy at the optimal range distance Z_{opt} = 21.3cm (Diesel engine at 1250 Hz).

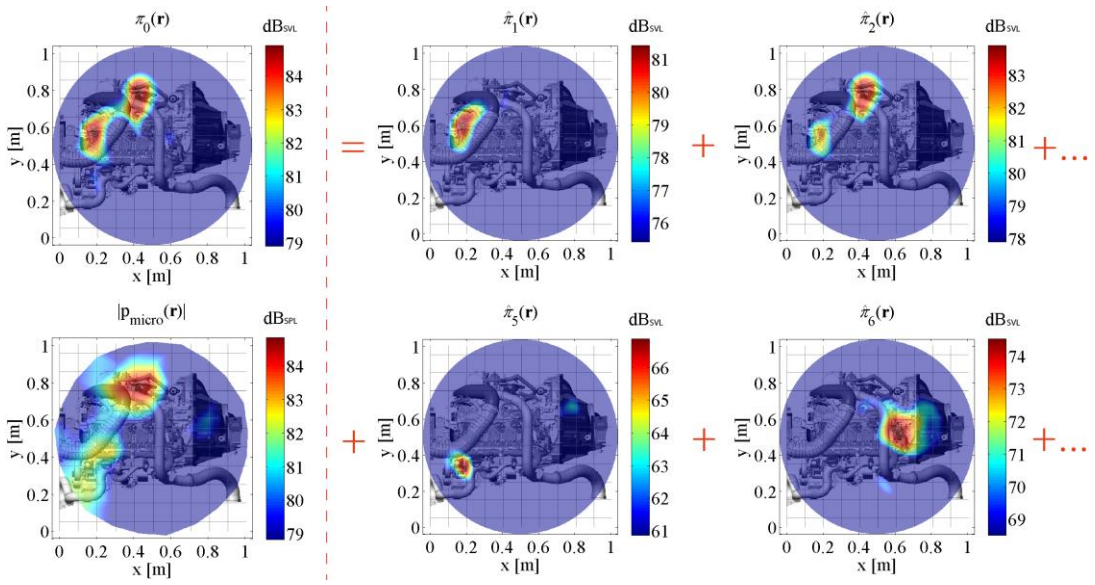

Fig. 7.22 Separated sources, $\hat{\pi}_i(\mathbf{r},\omega)$, i = 1,…,8, from the criterion of least spatial entropy at range distance Z = 9 cm (Diesel engine at 1250 Hz).

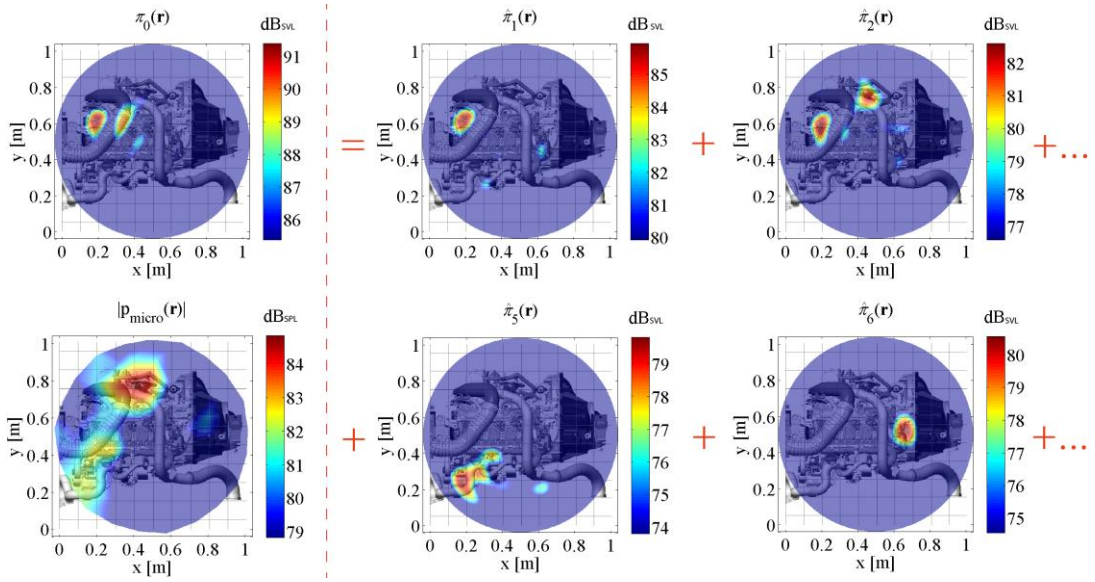

Fig. 7.23 Separated sources, $\hat{\pi}_i(\mathbf{r},\omega)$, i = 1,…,8, from the criterion of least spatial entropy at range distance Z = 27 cm (Diesel engine at 1250 Hz).

	s_1	s_2	s_3	s_4
Z = 9 cm	(17.8, 60)	(42.8, 80.4)	(18.8, 33.4)	(61.2, 53.2)
Z = 18 cm	(20.4, 62)	(44, 76.8)	(20.6, 27.6)	(66.6, 54.4)
Z = 21.3 cm	(20, 61.4)	(44.4, 76)	(20.2, 27.6)	(67.6, 55.6)
Z = 27 cm	(19.2, 61.6)	(46, 73.8)	(18, 24.6)	(66.4, 52)

Tab. 7.3 Coordinates (x, y) (in centimeter) of the highest peaks of four separated sources versus range different distances Z (Diesel engine at 1250 Hz).

7.3 Size of the aperture function

In the present dissertation, the aperture function is selected as a circular Hanning window, which is one candidate among many others [157]. The proposed algorithms for BSS can work with other types of windows, but this is out of the research scope of the dissertation. In this section, the size of aperture function is studied in short.

For the laboratory experiments, the radius of the aperture function was selected as R = 21 cm and R = 35 cm in Chapter 6. According to the separation results introduced in Chapter 6, this parameter has little effect on the separation performance of the proposed algorithms, except for mild distortion in shape and negligible bias in amplitude. For the Diesel engine example, the radius of the aperture function was R = 54 cm; separation results with radius set to R_1 = 45 cm and R_2 = 65 cm are depicted in Figs. 7.24 and 7.25, respectively. Except for the third source – $\hat{\pi}_5(\mathbf{r})$, the other three primary sources are little affected by the size of the aperture function. The reason is simple: the third source is close to the boundary of the aperture function as shown in Fig. 7.25. The coordinates of the hottest points of the four sources confirms the aforementioned observations. That is to say, CLSE is robust to the size of aperture function in such an industrial example. Based on the above discussion, the size of aperture function should be naturally chosen so as to enclose all sources to be separated.

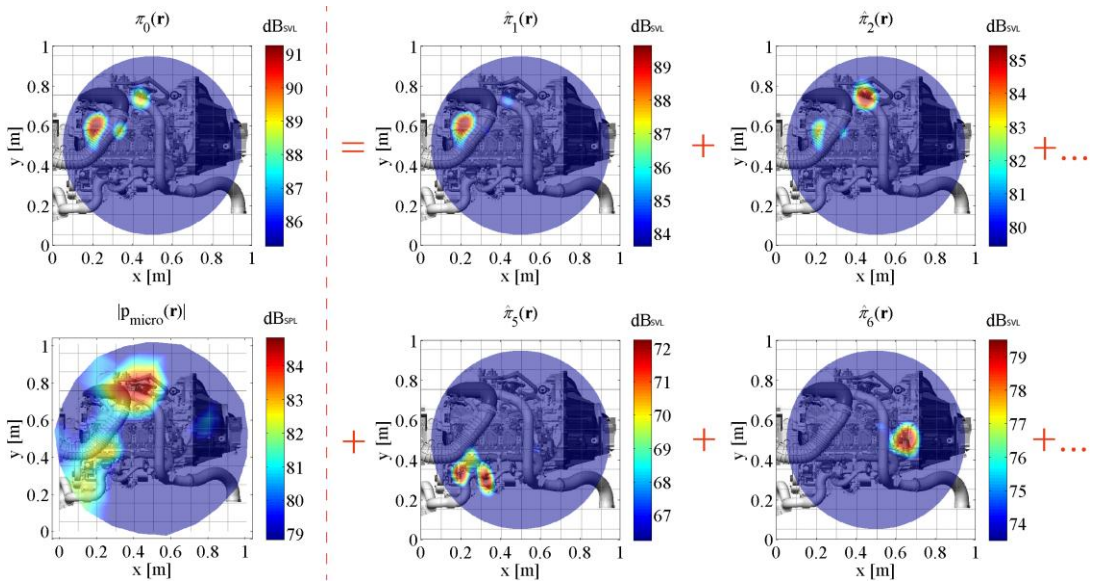

Fig. 7.24 Separated sources, $\hat{\pi}_i(\mathbf{r},\omega)$, i = 1,…,8, from the criterion of least spatial entropy with an aperture function of radius R = 45 cm (Diesel engine at 1250 Hz).

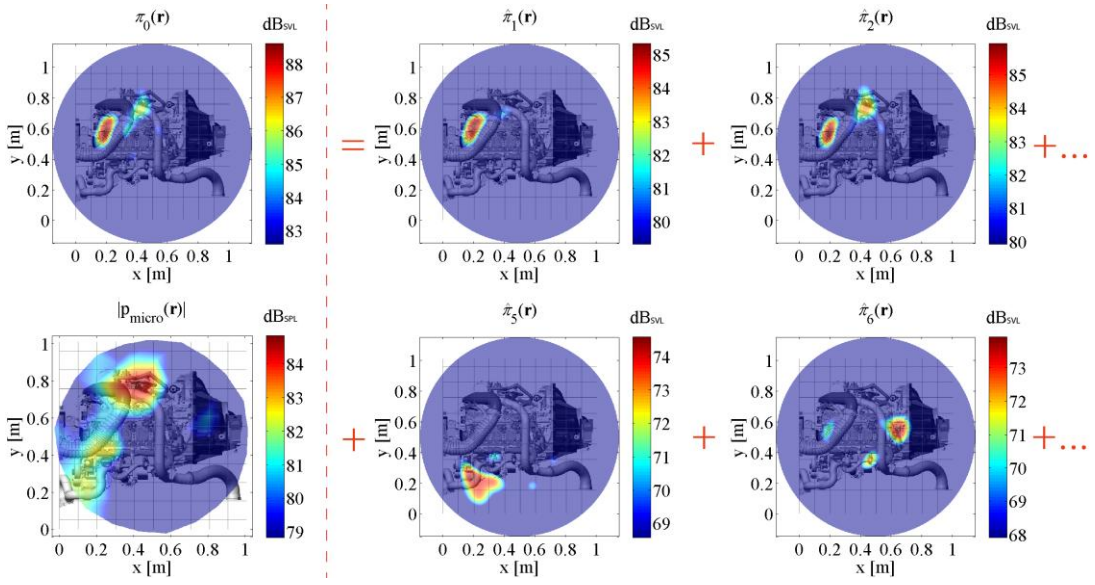

Fig. 7.25 Separated sources, $\hat{\pi}_i(\mathbf{r},\omega)$, i = 1,…,8, from the criterion of least spatial entropy with an aperture function of radius R = 65 cm (Diesel engine at 1250 Hz).

	s_1	s_2	s_3	s_4
R = 45 cm	(21.4, 60.4)	(44.8, 73.8)	(20, 32.8)	(65.4, 50)
R = 54 cm	(20.4, 62)	(44, 76.8)	(20.6, 27.6)	(66.6, 54.4)
R = 65 cm	(20.4, 61.4)	(43.6, 73.6)	(19.6, 23.6)	(65.2, 54.8)

Tab. 7.4 Coordinates (x, y) in centimeters of the highest peaks of four separated sources versus the radius R of aperture function (Diesel engine at 1250 Hz).

Conclusions

Blind separation of sound sources evidences a number of subtleties. One major issue is that it involves a spatial dimension not found in classical BSS. This is both a difficulty (since it corresponds to an additional convolution over space), and an advantage since it offers the possibility of defining new separation criteria. This dissertation demonstrates that a unique solution to blind source separation is found by forcing the spatial disjointing or the spatial compactness of the separated sources, two assumptions which are fairly realistic in many situations. The search for spatially disjoint sources has led to the so-called criterion of "spatial decorrelation" and the search of spatially compact sources to the so-called criteria of "least-spatial variance" and of "least-spatial entropy". The proposed solutions are fully original and have been found successful in situations where classical BSS methods fall short.

In the first part, based on the backpropagation from the measurements to the source domain of interest, the reason is firstly explained in clear: missing the rotation matrix, why the virtual sources from only statistical decorrelation do not equal the actual sources.

Inspired by the statistical decorrelation, the criterion of spatial decorrelation is proposed in Chapter 3 and applied to blindly separate sound sources which are disjoint in space. Meanwhile, the answer to the question proposed 28 years ago – "when do the virtual sources equal the actual sources?" – is found in success.

The principle of least spatial complexity is introduced in Chapter 4 under the assumption of spatial compactness. The two variables – the *spatial variance* and the *spatial entropy* involved in principle of least spatial complexity, both accurately measure the spatial compactness in quantity. Furthermore, the corresponding cost functions deduced from the two spatial criteria are convex, which apparently simplifies the optimization process and guarantees that the local minimum is equal to the total minimum.

It is still an open problem to determine the number of sources from measurements in the field of array signal processing. The contribution to the open problem is illustrated at the end of Chapter 5. That is the criteria of least spatial entropy leads to the proposal of a new method, the *entropic L-curve*, to determine the number of active sources by searching for the balance between spatial and statistical entropies. To be safe, it is suggested to combine the entropic L-curve with other classic methods to determine the active sources together.

In the second part, the proposed algorithms have been validated on laboratory, numerical, and industrial data. They also apply indifferently to the near-field and to the far-field. It can efficiently separate compact sources even with rather complex shapes, such as quadrupoles. Furthermore, it can separate very low sources whose amplitude is 40 dB less than the loudest source. Afterwards, the robustness of the algorithms to the experimental parameters is fully demonstrated on the laboratory and the industrial data.

One definite advantage of the approach in the dissertation is that it can separate sound sources under weak statistical assumptions (only decorrelation is actually needed), including the case of stationary and Gaussian sources (which classical BSS, for instance SOBI, has strong difficulties to deal with no matter in the near-field or in the far-field).

Four points are to be emphasized again. First, the virtual sources obtained from simple statistical decorrelation coincide with the actual sources if and only if the sources are also spatially uncorrelated. Second, the criterion of least spatial entropy is the first choice when searching for compact sources due to its accuracy, although it involves more computational demand than the criteria of spatial decorrelation and of least spatial variance. Third, the separation performance (whatever the criterion) strongly relies on the ability to first backpropagate the measurements from the array to the source domain. At last, the separation seems robust to the determination of the number of active sound sources, the range distance between the reconstruction plane and the array of microphones, and the size of the aperture function.

The end of a PhD thesis by no means indicates the end of a research topic. Many

issues on such topic are not clear enough or even still unknown. One interesting question for future research is whether the proposed criteria could be extended to separate (partially) coherent sound sources, for instance to account for the presence of wave reflections. With the help of the optimal spatial basis, it is very easy to determine the backpropagation involving contributions from actual sources and their reflected ones. Here, the actual source and its reflections are coherent to each other. The violation of the ordinary assumption of incoherence implies that much stricter spatial criterion is needed, if the principle of least spatial complexity can be applied to blindly separate the coherent sound sources.

Another point of interest is whether the proposed methodology could be used to separate sound sources radiated from rotating machines, e.g. the wind turbine. Due to the effect of rotation, the components which correspond to the sound origins are not stationary any more. That is to say, the backpropagation in the source domain of interest is non-stationary. The proposed methodology might be employed to separate such kind of non-stationary sources, if the measurements from an array of microphones can be stationarized first with success.

Acknowledgement

First of all, I would like to sincerely appreciate my supervisor – Prof. Jérôme ANTONI from the bottom of my heart. Without his guide and support, the dissertation could not be finished on time. No laudatory words could completely and precisely describe how excellent he is in my eyes. In 2011, Prof. J Jérôme ANTONI organized a summer school – 'Non-Stationary signal processing and inverse problem in acoustics and vibrations' and then made a concise prologue where he cited Benjamin Franklin's aphorism on teaching. Inspired by such a thought-provoking aphorism, I suddenly recognized that he belongs to the supervisors who have been highly praised for more than 1200 years in China. The criterion of evaluation has been clearly defined 1210 years ago by Han Yu, one of the greatest Chinese writers. "The supervisor is the man who teaches you, guides you, and explains you (Chinese version: 师者，所以授业传道解惑也。)".

I'd like to acknowledge my family in deep. Without the supports from my parents, I would have no chance to study in France; without the encouragements from my wife, I would not have so much power to conquer the problems one by one; without the love from my son, I would not begin to really understand what the sense of my life is.

I miss the times with Dr. Erliang ZHANG and his wife – Dr. HO Hsin Shen. At the beginning of my life in France, they helped me a lot and made me suit the life in broad quickly. Moreover, Dr. ZHANG provided me with many valuable scientific suggestions, mainly on the reconstruction of sound field and blind source separation. I also recall Jean-Marc's coffee which fully contains the happiness and friendship.

I appreciate very much the assistance in the experiments from Félix FOUCART and Jean-Marc GHERBEZZA, Laboratory Roberval UMR7337, University of Technology of Compiègne (UTC).

I am grateful to the encouragements from Prof. Charles PEZERAT, LAUM, Université du Maine. I enjoy very much the lectures of Prof. Johan SCHOUKENS,

Prof. Rik PINTELON, Prof. Yves ROLAIN and Prof. Gerd VANDERSTEEN in Dept. ELEC at Vrije Universiteit Brussel (VUB). Their experiences and actions tell me to be positive to failure in life, work hard and get up earlier. I admire the guidelines from Prof. Dr.-Ing. Walter KELLERMANN for the short scientific visit and appreciate the hosting of Institute of Multimedia Communications and Signal Processing, University of Erlangen-Nuremberg, Germany.

I appreciate all my colleagues and classmates at INSA de Lyon and at UTC. I thank their helps sincerely when I fell into troubles. I would not list all the names, because I would not like to forget anybody. I would like to acknowledge all people known and unknown, who once helped me in the past three years and half, too.

Finally, I am grateful to the China Scholarship Council (CSC) for the doctoral research funding in France. The support from Action – TU 1105 'NVH Analysis Techniques for Design and Optimization of Hybrid and Electric vehicles' under the framework of European Cooperation in Science and Technology (COST) is appreciated in deep. Meanwhile, I also thank the support from the Acoustical Center of Lyon (CeLyA).

References

[1] F. Fahy, J. Walker. Advanced Applications in Acoustics, Noise and Vibration. London : Spon Press, 2004, 656 p.

[2] Q. Leclère. Étude et développement de la mesure indirecte d'efforts: Application à l'identification des sources internes d'un moteur Diesel (Research and development of the indirect measure of efforts: application on the identification of internal sources of a Diesel engine). Doctoral dissertation from GMC. Lyon : University of Lyon, 2003, 165 p (in French).

[3] C. Pezerat, Q. Leclère, N. Totaro et al. Identification of vibration excitations from acoustic measurements using near field acoustic holography and the force analysis technique. Sound and Vibration, 2009, vol. 326, pp. 540-556.

[4] Q. Leclère, C. Pezerat. Vibration source identification using corrected finite difference schemes. Sound and Vibration, 2012, vol. 331, pp. 1366-1377.

[5] P. Sijtsma. Using phased array beamforming to identify broadband noise sources in a turbofan engine. Key Note Presentaions of 12th CEAS-ASC Workshop Turbomachinery Broadband Noise, 23-24 October 2008, Bilbao, Spain,.

[6] Acoustic camera. Available on : http://www.acoustic-camera.com/ (Consult at 13:24, on 16.11.2013).

[7] B. Lafon. Séparation de sources couplée aux techniques d'imagerie acoustique: Application au rayonnement de moteurs Diesel (Separation of sources coupled with techniques of acoustic imaging: application on the influence of Diesel engines). Doctoral dissertation from Laboratory Roberval UMR7337. Compiègne : University of Technology of Compiègne, 2009, 198 p (in French).

[8] B. Lafon, J. Antoni, M. Sidahmed et al. The concept of cyclic sound intensity and its application to acoustical imaging. Sound and Vibration, 2011, vol. 330, pp. 2107-2121.

[9] J. Billingsley, R. Kinns. The acoustic telescope. Sound and Vibration, 1976, vol. 48, pp. 485-510.

[10] U. Michel. History of acoustic beamforming. The 1st Berlin Beamforming Conference (BeBeC), 2006, Berlin, Germany, pp. 1-17.

[11] J. D. Maynard, E. G. Williams, Y. Lee. Nearfield acoustic holography: Ⅰ. Theory of generalized holography and the development of NAH. Acoustical Society of America, 1985, vol. 78, pp. 1395-1413.

[12] E. G. Williams. Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography. New York : Academic Press, 1999, 306 p.

[13] J. Hald. Basic theory and properties of statistically optimized near-field acoustical holography. Acoustical Society of America, 2009, vol. 125, pp. 2105-2120.

[14] F. Jacobsen, V. Jaud. Statistically optimized near field acoustic holography using an array of pressure-velocity probes. Acoustical Society of America, 2007, vol. 121, pp. 1550-1558.

[15] S. F. Wu. On reconstruction of acoustic pressure fields using the Helmholtz equation least squares method. Acoustical Society of America, 2000, vol. 107, pp. 2511-2522.
[16] A. Sarkissian. Extension of measurement surface in near-field acoustic holography. Acoustical Society of America, 2004, vol. 115, pp. 1593-1596.
[17] Q. Leclère. Acoustic imaging using under-determined inverse approaches: Frequency limitations and optimal regularization. Sound and Vibration, 2009, vol. 321, pp. 605-619.
[18] A. Pereira. Acoustic imaging in enclosed spaces. Doctoral dissertation from GMC. Lyon : University of Lyon, 2013, 205 p.
[19] A. Schuhmacher, J. Hald, K. B. Rasmussen et al. Sound source reconstruction using inverse boundary element calculations. Acoustical Society of America, 2003, vol. 113, pp. 114-127.
[20] A. Oey, H. W. Jang, J. G. Ih. Effect of sensor proximity over the non-conformal hologram plane in the near-field acoustical holography based on the inverse boundary element method. Sound and Vibration, 2010, vol. 329, pp. 2083-2098.
[21] M. B. S. Magalhães, R. A. Tenenbaum. Sound sources reconstruction techniques: A review of their evolution and new trends. Acta Acustica united with Acustica, 2004, vol. 90, pp. 199-220.
[22] S. F. Wu. Methods for reconstructing acoustic quantities based on acoustic pressure measurements. Acoustical Society of America, 2008, vol. 124, pp. 2680-2697.
[23] E. J. Candes, J. Romberg, T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on Information Theory, 2006, vol. 52, pp. 489-509.
[24] E. J. Candes, J. Romberg, T. Tao. Stable signal recovery from incomplete and inaccurate measurements. Communications on Pure and Applied Mathematics, 2006, vol. 59, pp. 1207-1223.
[25] D. L. Donoho, Compressed sensing, Journal of IEEE Transactions on Information Theory 52 (2006) 1289-1306.
[26] G. Chardon, L. Daudet, A. Peillot et al. Near-field acoustic holography using sparse regularization and compressive sampling principles. Acoustical Society of America, 2012, vol. 132, pp. 1521-1534.
[27] J. Antoni. A Bayesian approach to sound source reconstruction: Optimal basis, regularization and focusing. Acoustical Society of America, 2012, vol. 131, pp. 2873-2890.
[28] Y. H. Kim. Can we hear the shape of a noise source? The 18th International Congress on Acoustics (ICA 2004), 2004, Kyoto, Japan, pp. 3357-3370.
[29] E. F. Grande. Near-field acoustic holography with sound pressure and particle velocity measurements. Doctoral dissertation from Department of Electrical Engineering. Kongens Lyngby : Technical University of Denmark, 2012, 200 p.
[30] N. Chu. Bayesian approach in acoustic source localization and imaging. Doctoral dissertation from L2S. Paris : University of Paris 11, 2013, 227 p.
[31] P. Comon. Independent component analysis: A new concept? Signal Processing,

1994, vol. 36, pp. 287-314.
[32] J. F. Cardoso. Blind signal Separation: Statistical principles. Proceeding of the IEEE, 1998, vol. 86, pp. 2009-2025.
[33] A. J. Bell, T. J. Sejnowski. An information-maximisation approach to blind separation and blind deconvolution. Neural Computation, 1995, vol. 7, pp. 1004-1034.
[34] T. W. Lee, M. Girolami, A. J. Bell et al. A unifying information-theoretic framework for independent component analysis. Computers & Mathematics with Applications, 2000, vol. 39, pp. 1-21.
[35] X. R. Cao, R. W. Liu. General approach to blind source separation. IEEE Transactions on Signal Processing, 1996, vol. 44, pp. 562-571.
[36] A. Hyvärinen, E. Oja. Independent component analysis: Algorithms and application. Neural Networks, 2000, vol. 13, pp. 411-430.
[37] K. E. Hild II, D. Erdogmus, J. C. Principe. An analysis of entropy estimators for blind source separation. Signal Processing, 2006, vol. 86, pp. 181-194.
[38] L. F. Guo, M. Garland. The use of entropy minimization for the solution of blind source separation problems in image analysis. Pattern Recognition, 2006, vol. 39, pp. 1066-1073.
[39] S. Shwartz, Y. Y. Schechner, M. Zibulevsky. Blind separation of convolutive image mixtures. Neurocomputing, 2008, vol. 71, pp. 2164-2179.
[40] X. L. Li, T. Adali. Blind separation of noncircular correlated sources using Gaussian entropy rate. IEEE Transactions on Signal Processing, 2011, vol. 59, pp. 2969-2975.
[41] T. Pham. Blind separation of instantaneous mixture of sources via the Gaussian mutual information criterion. Signal Processing, 2001, vol. 81, pp. 855-870.
[42] A. Hyvärinen, J. Karhunen, E. Oja. Independent Component Analysis. New York : Wiley, 2001, pp. 147-163.
[43] A. Cichocki, S. Amari. Adaptive blind signal and image processing – Learning algorithms and applications. Chichester : Wiley, 2002, 586 p.
[44] P. Comon, C. Jutten. Handbook of Blind Source Separation: Independent Component Analysis and Application. Oxford : Academic Press, 2010, p.
[45] J. Hérault, C. Jutten. Space or time adaptive signal processing by neural networks models. International Conference on Neural Networks for Computing, Snowbird, 1986, Utah, USA, pp. 206-211.
[46] S. Amari, A. Cichocki. Adaptive blind signal processing-neural network approaches. Proceedings of the IEEE, 1998, vol. 86, pp. 2026-2048.
[47] J. F. Cardoso, A. Souloumiac. Blind beamforming for non Gaussian signals. Radar and Signal Processing, IEEE Proceedings F, 1993, vol. 140, pp. 362-370.
[48] A. Belouchrani, K. A. Meraim, J. F. Cardoso et al. A blind source separation technique using second-order statistics. IEEE Transactions on Signal Processing, 1997, vol. 45, pp. 434-444.
[49] R. Gribonval, S. Lesage. A survey of sparse components analysis for blind source separation: Principle, perspectives and new challenges. European Symposium on Artificial Neural Networks (ESANN 2006 Proceedings), 2006, Bruges, Belgium,

pp. 323-330.
[50] A. L. Casanovas, G. Monaci, P. Vandergheynst et al. Blind Audiovisual Source Separation Based on Sparse Redundant Representations. IEEE Transactions on Multimedia, 2010, vol. 12, pp. 358-371.
[51] M. Kleinsteuber, S. Hao. Blind Source Separation with Compressively Sensed Linear Mixtures. IEEE Signal Processing Letters, 2012, vol. 19, pp. 107-110.
[52] J. S. Bendat, A. G. Piersol. Random Data: Analysis and Measurement Procedures. New York : Wiley, 1971, 407 p.
[53] D. Hallman, J. S. Bolton. Multi-reference near-field acoustical holography. Proceedings of Inter-noise'92, July 1992, Toronto, Ontario, Canada, pp. 1165-1170.
[54] D. L. Hallman, J. S. Bolton. Comparison of multi-reference nearfield acoustical holography procedures. Proceedings of National Conference on Noise Control Engineering, May 1994, Fort Lauderdale, FL, USA, pp. 929-934.
[55] J. F. Li, J. C. Pascal, C. Carles. Reconstruction of partially coherent sources by use of principal component analysis. Proceedings of Inter-noise'95, July 1995, Newport Beach, CA, USA, pp. 1355-1358.
[56] R. J. Ruhala, C. B. Burroughs. Separation of leading edge, trailing edge, and sidewall noise sources from rolling tires. Proceedings of NOISE-CON 98, April 1998, Ypsilanti, MI, USA, pp. 109-114.
[57] M. A. Tomlinson. Partial source discrimination in near field acoustic holography. Applied Acoustics, 1999, vol. 57, pp. 243-261.
[58] S. M. Price, R. J. Bernhard. Virtual coherence: A digital signal processing technique for incoherent source identification. Proceedings of the 4th International Modal Analysis Conference, vol. 2, 1986, Los Angeles, CA, USA, pp. 1256-1262.
[59] K. U. Nam, Y. H. Kim. Visualization of multiple incoherent sources by the backward prediction of near-field acoustic holography. Acoustical Society of America, 2001, vol. 109, pp. 1808-1816.
[60] Y. J. Kim, J. S. Bolton, H. S. Kwon. Partial sound field decomposition in multireference near-field acoustical holography by using optimally located virtual references. Acoustical Society of America, 2004, vol. 115, pp. 1641-1652.
[61] M. Y. Lee, J. S. Bolton. Scan-based near-field acoustical holography and partial field decomposition in the reference of noise and source level variation. Acoustical Society of America, 2006, vol. 119, pp. 382-393.
[62] K. U. Nam, Y. H. Kim. A partial field decomposition algorithm and its examples for near-field acoustic holography. Acoustical Society of America, 2004. vol. 116, pp. 172-185.
[63] Z. M. Zhong, J. Chen, P. Zhong et al. Application of the blind source separation method to feature extraction of machine sound signals. Advanced Manufacturing Technology, 2006, vol. 28, pp. 855-862.
[64] K. Teramoto, M. T. I. Khan. Real-time acoustic blind signal separation system based on the spatio-temporal gradient analysis. The International Conference of Acoustics 08 Paris, June 2008, Paris, France, pp. 111-116.

[65] R. Aichner, H. Buchner, F. Yan et al. A real-time blind source separation scheme and its application to reverberant and noisy acoustic environments. Signal Processing, 2006, vol. 86, pp. 1260-1277.
[66] M. Zibulevsky, B. A. Pearlmutter. Blind source separation by sparse decomposition in a signal dictionary. Neural Computation, 2001, vol. 13, pp. 863-882.
[67] E. Vincent, R. Gribonval, C. Févotte. Performance Measurement in blind audio source separation. IEEE Transactions on Speech and Audio Processing, 2006, vol. 14, pp. 1462-1469.
[68] B. H. Wu, G. P. Too, S. Lee. Audio signal separation via a combination procedure of time-reversal and deconvolution process. Mechanical Systems and Signal Processing, 2010, vol. 24, pp. 1431-1443.
[69] E. Zhang, J. Antoni, B. Dong et al. Bayesian space-frequency separation of wide-band sound sources by a hierarchical approach. Acoustical Society of America, 2012, vol. 132, pp. 3240-3250.
[70] D. R. Brillinger. Time Series: Data Analysis and Theory. Philadelphia : SIAM, 2001, 540 p.
[71] J. Antoni, S. Chauhan. A study and extension of second-order blind source separation to operational modal analysis. Sound and Vibration, 2013, vol. 332, pp. 1079-1106.
[72] P. A. Nelson, S. H. Yoon. Estimation of acoustic source strength by inverse methods: Part I, Conditioning of the inverse problem. Sound and Vibration, 2000, vol. 233, pp. 643-668.
[73] S. H. Yoon, P. A. Nelson. Estimation of acoustic source strength by inverse methods: Part II, Experimental investigation of methods for choosing regularization parameters. Sound and Vibration, 2000, vol. 233, pp. 669-705.
[74] P. A. Nelson. A review of some inverse problems in acoustics. Acoustics and Vibration, 2001, vol. 6, pp. 118-134.
[75] Y. Kim, P. A. Nelson. Spatial resolution limits for the reconstruction of acoustic source strength by inverse methods. Sound and Vibration, 2003, vol. 265, pp. 583-608.
[76] Y. Kim, P. A. Nelson. Optimal regularization for acoustic source reconstruction by inverse methods. Sound and Vibration, 2004, vol. 275, pp. 463-487.
[77] E. G. Williams. Regularization methods for near-field acoustical holography. Acoustical Society of America, 2001, vol. 110, pp. 1976-1988.
[78] A. Tarantola. Inverse Problem Theory and Methods for Model Parameter Estimation. Philadelphia : SIAM, 2005, 352 p.
[79] J. Idier. Bayesian Approach to Inverse Problems. London : WILEY, 2008, 392 p.
[80] A. Kirsch. An introduction to the Mathematical Theory of Inverse Problems. 2nd Ed. Heidelberg : Springer, 2011, 334 p.
[81] J. Hald. STFT – A unique technique for scan-based near-field acoustic holography without restrictions on coherence. No. 1. B & K Technical Review, 1989, 56 p.
[82] S. H. Yoon, P. A. Nelson. A method for the efficient construction of acoustic pressure cross-spectral matrices. Sound and Vibration, 2000, vol. 233, pp.

897-920.
[83] Y. Kim, P.A. Nelson. Estimations of acoustic source strength within a cylindrical duct by inverse problems. Sound and Vibration, 2004, vol. 275, pp. 391-413.
[84] P.A. Nelson, J. F. W. Rose. The time domain response of some systems for sound reproduction. Sound and Vibration, 2006, vol. 296, pp. 461-493.
[85] K. R. Holland, P. A. Nelson. An experimental comparison of the focused beamformer and the inverse method for the characterization of acoustic sources in ideal and non-ideal acoustic environments. Sound and Vibration, 2012, vol. 331, pp. 4425-4437.
[86] K. R. Holland, P. A. Nelson. The application of inverse methods to spatially-distributed acoustic sources. Sound and Vibration, 2013, vol. 332, pp. 5727-5747.
[87] J. B. Fahnline, G. H. Koopmann. A numerical solution for the general radiation problem based on the combined methods of superposition and singular-value decomposition. Acoustical Society of America, 1991, vol. 90, pp. 2808-2819.
[88] A. D. Polyanin, A. V. Manzhirov. Handbook of integral equations. 2nd Ed. London : CRC Press LLC, 2008, pp. 573-706.
[89] R. Pintelon, J. Schoukens. System identification: A frequency domain approach. 2nd Ed. New York : Wiley-IEEE Press, 2012, 788 p.
[90] S. Choi, A. Cichocki, H. M. Park et al. Blind source separation and independent component analysis : A Review. Neural Information Processing-Letters and Reviews, 2005, vol. 6, pp. 1-57.
[91] J. Bobin, J. L. Starck, Y. Moudden et al. Blind source separation: The sparsity revolution. Advanced in Imaging and Electron Physics, 2008, vol. 152. pp. 221-302.
[92] F. J. Theis, A. Jung, C. G. Puntonet et al. Linear geometric ICA: Fundamentals and algorithms. Neural Computation, 2003, vol. 15, pp. 419-439.
[93] C. Bi, X. Chen, R. Zhou et al. Landweber iterative regularization for nearfield acoustic holography. Chinese Science Bulletin, 2006, vol. 51, pp. 1374-1380.
[94] J. C. Pascal, S. Paillasseur, J. H. Thomas et al. Patch near-field acoustic holography: Regularized extension and statistically optimized methods. Acoustical Society of America, 2009, vol. 126, pp. 1264-1268.
[95] S. Paillasseur, J. H. Thomas, J. C. Pascal. Regularization for improving the deconvolution in real-time near-filed acoustic holography. Acoustical Society of America, 2011, vol. 129, pp. 3777-3787.
[96] Whitening transformation. Available on : http://en.wikipedia.org/wiki/Whitening_transformation, (Consult at 09:39, on 14.10.2013).
[97] F. J. Fahy. Sound Intensity. 2nd Ed. London : E & FN SPON, 1995, 320 p.
[98] L. L. Scharf. Statistical Signal Processing: Detection, Estimation, and Time Series Analysis. Amsterdam : Addision-Wesley, 1991, 524 p.
[99] M. E. Tipping, C. M. Bishop. Probabilistic principal component analysis. Royal Statistical Society, Series B : Statistical Methodology, 1999, vol. 61, pp. 611-622.

[100] J. F. Cardoso. Available on : http://sig.enst.fr/~cardoso/stuff.html, (Consult at 13:26, on 14.10.2013).
[101] Entropy. Available on : http://en.wikipedia.org/wiki/Entropy, (Consult at 15:03, on 23.10.2013).
[102] E. Shannon. A mathematical theory of communication. The Bell System Technical Journal, 1948, vol. 27, pp. 379-423, 623-656.
[103] E. T. Jaynes. Information theory and statistical mechanics. Physical Review, 1957, vol. 106, pp. 620-630.
[104] A. G. Wilson. Entropy in Urban and Regional Modelling. London : Pion Press, 1970, 166 p.
[105] N. G. Roegen. The Entropy Law and the Economic Process. Cambridge : Harvard Univrsity Press, 1971, p.
[106] M. Batty. Spatial entropy. Geographical Analysis, 1974, vol. 6, pp. 1-31.
[107] M. Batty. Entropy in spatial aggregation. Geographical Analysis, 1976, vol. 8, pp. 1-21.
[108] H. Krim, M. Viberg. Two decades of array signal processing research: The parametric approach. IEEE Signal Processing Magazine, 1996, vol. 13, pp. 67-96.
[109] J. H. Manton. On the role of differential geometry in signal processing. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2005), vol. 5, 18-23 March 205, Philadelphia, Pennsylvania, USA, pp. 1021-1024.
[110] J. H. Manton. Optimization algorithms exploiting unitary constraints. IEEE Transactions on Signal Processing, 2002, vol. 50, pp. 635-650.
[111] T. E. Abrudan, J. Eriksson, V. Koivunen. Optimization under unitary matrix constraint using approximate matrix exponential. Conference Record of the Thirty-Ninth Asilomar Conference On Signals, Systems and Computers (ACSSC 2005), 2005, Pacific Grove, CA, USA, pp. 242-246.
[112] T. Abrudan, J. Eriksson, V. Koivunen. Efficient Riemannian algorithms for optimization under unitary matrix constraint. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008), 2008, Las Vegas, NV, USA, pp. 2353-2356.
[113] T. E. Abrudan, J. Eriksson, V. Koivunen. Steepest descent algorithms for optimization under unitary matrix constraint. IEEE Transactions on Signal Processing, 2008, vol. 56, pp. 1134-1147.
[114] T. Abrudan, J. Eriksson, V. Koivunen. Conjugate gradient algorithm for optimization under unitary matrix constraint. Signal Processing, 2009, vol. 89, pp. 1704-1714.
[115] T. Abrudan. Advanced optimization algorithms for sensor arrays and multi-antenna communications. Doctoral dissertation from Faculty of Electronics, Communications and Automation. Espoo : Helsinki University of Technology, 2008, pp. 33-50.
[116] T. Abrudan, J. Eriksson, V. Koivunen. Efficient line search methods for Riemannian optimization under unitary matrix constraint. Conference Record of

the Forty-First Asilomar Conference On Signals, Systems and Computers (ACSSC 2007), 2007, Pacific Grove, CA, USA, pp. 671-675.
[117] D. H. Brandwood. A complex gradient operator and its application in adaptive array theory. IEE Proceedings F: Communications, Radar and Signal Processing, 1983, vol. 130, pp. 11-16.
[118] A. Hjørungnes, D. Gesbert. Introduction to complex-valued matrix differentiation. [Online]. Available on: http://www.unik.no/personer/arehj/publications/matrixdiff.pdf, (Consult at 16:13, on 23.10.2013).
[119] A. Hjørungnes, D. Gesbert. Complex-valued matrix differentiation: Techniques and key results. IEEE Transactions on Signal Processing, 2007, vol. 55, pp. 2740-2746.
[120] A. Hjørungnes, D. Gesbert, D. P. Palomar. Unified theory of complex-valued matrix differentiation. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2007), 2007, Honolulu, HI, USA, pp. 345-348.
[121] E. Ollila, V. Koivunen, H. V. Poor. Complex-valued signal processing – essential models, tools and statistics. Information Theory and Applications Workshop (ITA), 2011, La Jolla, CA, USA, pp. 1-10.
[122] A. Hjørungnes. Complex-Valued Matrix Derivatives: With Applications in Signal Processing and Communications. Cambridge : Cambridge Univeristy Press, 2011, 270 p.
[123] F. Fahy, P. Gardonio. Sound and Structural Vibration: Radiation, Transmission and Response. 2nd Ed. Oxford : Academic Press, 2007, 656 p.
[124] D. B. Williams. Detection: Determining the number of sources. In : V. Madisetti, D. Williams Ed. Digital Signal Processing Handbook. London : CRC Press, 1999, pp. 1436-1445.
[125] M. S. Bartlett. A note on the multiplying factors for various χ^2 approximations. Royal Statistical Society, Series B: Methodological, 1954, vol. 16, pp. 296-298.
[126] D. N. Lawley. Tests of significance of the latent roots of the covariance and correlation matrices. Biometrika, 1956, vol. 43, pp. 128-136.
[127] D. B. Williams, D. H. Johnson. Using the sphericity test for source detection with narrow-band passive arrays. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1990, vol. 38, pp. 2008-2014.
[128] M. Wax, T. Kailath. Detection of signals by information theoretic criteria. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1985, vol. 33, pp. 387- 392.
[129] H. Akaike. Information theory and extension of the maximum likelihood principle. the Second International Symposium on Information Theory, 1973, Budapest, Hungary, pp. 267-281.
[130] H. Akaike. A new look at the statistical model identification. IEEE Transactions on Automatic Control, 1974, vol. 19, pp. 716-723.
[131] H. Akaike. Likelihood of a model and information criteria. Economics, 1981, vol. 16, pp. 3-14.
[132] G. Schwarz. Estimating the dimension of a model. Annals of Statistics, 1978, vol.

6, pp. 461-464.
[133] J. Rissanen. Modeling by shortest data description. Automatica, 1978, vol. 14, pp. 465-471.
[134] J. Rissanen. A universal priori for the integers and estimation by minimum description length. Annals of Statistics, 1983, vol. 11, pp. 416-431.
[135] L. C. Zhao, P. R. Krishnaiah, Z. D. Bai. Remarks on criteria for detection of number of signals. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1987, vol. 35, pp. 129-132.
[136] K. M. Wong, Q. T. Zhang, J. P. Reilly et al. On information theoretic criteria for determining the number of signals in high resolution array processing. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1990, vol. 38, pp. 1959-1971.
[137] W. Xu, M. Kaveh. Analysis of the performance and sensitivity of eigendecomposition-based detectors. IEEE Transactions on Signal Processing, 1995. vol. 43, pp. 1413-1426.
[138] A. P. Liavas, P. A. Regalia. On the behavior of information theoretic criteria for model order selection. IEEE Transactions on Signal Processing, 2001, vol. 49, pp. 1689-1695.
[139] E. Fishler, M. Grosmann, H. Messer. Detection of signals by information theoretic criteria: General asymptotic performance analysis. IEEE Transactions on Signal Processing, 2002, vol. 50, pp. 1027-1035.
[140] S. Kritchman, B. Nadler. Non-Parametric detection of the number of signals: Hypothesis testing and random matrix theory. IEEE Transactions on Signal Processing, 2009, vol. 57, pp. 3930-3941.
[141] B. Nadler. Nonparametric detection of signals by information theoretic criteria: Performance analysis and an improved estimator. IEEE Transactions on Signal Processing, 2010, vol. 58, pp. 2746-2756.
[142] L. C. Zhao, P. R. Krishnaiah, Z. D. Bai. On detection of the number of signals in presence of white noise. Multivariate Analysis, 1986, vol. 20, pp. 1-25.
[143] D. B. Williams. Comparison of AIC and MDL to the minimum probability of error criterion. IEEE Sixth SP Workshop on Statistical Signal and Array Processing, 1992, Victoria, BC, Canada, pp. 114-117.
[144] D. Otto, P. Sas, P. Van de Ponseels. Principal component analysis for noise source identification. Proceedings of the 6th IMAC, 1988, Florida, USA, pp. 1207-1214.
[145] I. T. Jolliffe. Principal Component Analysis. Berlin : Springer, 2002, 489 p.
[146] C. T. Latombe. Détection et caractérisation des signaux à plusieurs composants à partir de la matrice interspectrale (Detection and characterization of several-component signal from the interspectral matrix). Doctoral dissertation from University of Grenoble. Grenoble : University of Grenoble, 1982, 165 p (in French).
[147] M. S. Kompella, P. Davies, R. J. Bernhard et al. A technique to determine the number of incoherent sources contributing to the response of a system. Mechanical Systems and Signal Processing, 1994, vol. 8, pp. 363-380.

[148] P. C. Hansen. Analysis of discrete ill-posed problems by means of the L-curve. SIAM Review, 1992, vol. 34, pp. 561-580.
[149] P. C. Hansen, D. P. O'Leary. The use of the L-curve in the regularization of discrete ill-posed problems. SIAM Review, 1993, vol. 14, pp. 1487-1503.
[150] P. C. Hansen. Rank-Deficient and Discrete Ill-Posed Problem: Numerical Aspects of Linear Inversion. Philadelphia : SIAM, 1998, 263 p.
[151] S. Kullback. Information Theory and Statistics. New York : Dover Publications, 1997, 432 p.
[152] J. Hald. Array designs optimized for both low-frequency NAH and high-frequency Beamforming. The 33rd International Congress and Exposition on Noise Control Engineering, 2004, Prague, Czech Republic, pp. 1-8.
[153] J. A. Raff, R. D. Perry. A review of vehicle noise studies carried out at the institute of sound and vibration research with a reference to some recent research on petrol engine noise. Sound and Vibration, 1973, vol. 28, pp. 433-470.
[154] R. J. Alfredson. The partial coherence technique for source identification on a Diesel engine. Sound and Vibration, 1977, vol. 55, pp. 487-494.
[155] J. Hald. Acoustical testing of a diesel engine using STFT. Brüel & Kjær Application notes, 1987, 12 p.
[156] M. F. Albright. Conditioned source analysis, a technique for multiple input system identification with application to combustion energy separation in piston engines. SAE paper 951376, 1995, pp. 1165-1177.
[157] F. J. Harris. On the use of windows for harmonic analysis with the discrete Fourier transform. Proceedings of the IEEE, 1978, vol. 66, pp. 51-83.

Appendices

Appendix A

Bayesian regularization

As introduced in Ref [27], the mechanism of Bayesian regularization is as follows: the optimal regularization parameter, i.e. the noise-to-signal ratio $\eta^2=\beta^2/\alpha^2$ (where β^2 denotes the mean energy of noise from measurements and α^2 for the mean energy of sources), makes the best trade-off between an acceptable solution with highly physical meaning and a reasonable fitting of the measurements with respect to the actual sources.

The key point in Bayesian regularization is how to adjust the parameter η^2, according to the given source distribution and topology of an array of sensors. Let $[\beta^2, \alpha^2|\mathbf{p}]$ denote the posterior distribution of the unknown hyperparamters β^2 and α^2 with respect to measurements **p**. If the priori is assumed to be the uniform distribution before recording data, the posterior $[\beta^2, \alpha^2|\mathbf{p}]$ will be proportional to the likelihood $[\mathbf{p}|\beta^2, \alpha^2]$ in terms of Bayes rule. Therefore, the optimal hyperparamters β_0^2 and α_0^2 are the ones who maximize the likelihood function $[\mathbf{p}|\beta^2, \alpha^2]$. A practical algorithm, which is thoroughly introduced in Ref [27], is applied here. Consequently, a criterion derived from the likelihood function $[\mathbf{p}|\beta^2, \alpha^2]$, is

$$J(\eta^2)=\underbrace{\ln(\hat{\alpha}^2)}_{i}+\underset{ii}{\frac{I}{M}}+\underset{iii}{\ln(\eta^2)}, \tag{A.1}$$

where $\hat{\alpha}^2$ denotes the estimated mean energy of sources, M symbolizes the number of microphones embedded in the array, and I defines the mutual information between sources $s(\mathbf{r})$ and measurements **p** as

$$I=\sum_{m=1}^{M}\left(1+\frac{s_m^2}{\eta^2}\right), \tag{A.2}$$

where s_m denotes the singular values. The three items in the cost function Eq. (A.1) are: *i*) the logarithm of the estimated mean energy, which decreases with η^2 rising; *ii*) the mutual information involved in each microphone; *iii*) the penalty part, respectively. Thus the optimal regularization parameter η_0^2, corresponding to the minimum of Eq. (A.1), makes a balance between the least energy and maximum transfer of information which explains the measurements.

At last, three highlights of Bayesian regularization are underlined here:

1) The cost function $J(\eta^2)$ (i.e. Eq. (A.1)) is convex, which guarantees the existence and uniqueness of the regularization parameter η_0^2. That means it is convenient to adjust η^2 in the theoretical framework of Bayesian regularization.

2) The optimization of the cost function $J(\eta^2)$ is very efficient, as only one "inversion" operation is required in the first item of Eq. (A.1). Therefore, Bayesian regularization obtains the regularization parameter η_0^2 much easier than some classic regularization methods do.

3) It returns the optimal regularization parameter η_0^2, and the mean energies β^2 and α^2 simultaneously, which provides the following reconstruction of the source filed with many benefits.

Appendix B

Optimization of CLSV in Stiefiel manifold

Based on the cost function of the missing matrix **V** from the criterion of least spatial variance (see Eq. (3.20)), the corresponding Euclidean gradient $\mathbf{E}_\Sigma$ is defined as

$$\mathbf{E}_\Sigma = \frac{\partial \Sigma(\mathbf{V})}{\partial \mathbf{V}^H}. \tag{B.1}$$

To highlight the relationship between the cost function $\Sigma(\mathbf{V})$ and the variable **V**, two intermediate variables are introduced,

$$F_l(\mathbf{v}_i) = \mathbf{e}_l^T \mathbf{\Phi U D V}^H \mathbf{e}_i \mathbf{e}_i^T \mathbf{V D U}^H \mathbf{\Phi}^H \mathbf{e}_l = \mathbf{q}_l^H \mathbf{v}_i^H \mathbf{v}_i \mathbf{q}_l, \tag{B.2}$$

and

$$\mathbf{q}_l = \mathbf{D U}^H \mathbf{\Phi}^H \mathbf{e}_l, \tag{B.3}$$

where $\mathbf{v}_i$ is the *i*-th row of matrix **V**. Substituting Eqs. (B.2) and (B.3) into Eqs. (3.20) and (3.21), the cost function of CLSV can be rewritten as

$$\Sigma(\mathbf{V}) = \sum_{i=1}^{N_s} \sum_{l=1}^{N} F_l(\mathbf{v}_i) |\mathbf{r}_l - \mathbf{r}_{0i}|^2 \Delta\Gamma(\mathbf{r}_l), \tag{B.4}$$

with

$$\mathbf{r}_{0i} = \frac{\sum_{l=1}^{N} F_l(\mathbf{v}_i) \mathbf{r}_l \Delta\Gamma(\mathbf{r}_l)}{Tr\{\mathbf{\Gamma \Phi U D V}^H \mathbf{e}_i \mathbf{e}_i^T \mathbf{V D U}^H \mathbf{\Phi}^H\}}. \tag{B.5}$$

Consequently, substituting Eqs. (B.4) and (B.5) into Eq. (B.1), the (*m*, *k*)-th element of the Euclidean gradient $\mathbf{E}_\Sigma$ can be formulated as

$$(\mathbf{E}_\Sigma)_{mk} = \frac{\partial \Sigma(\mathbf{V})}{\partial v_{km}^*} = \sum_{i=1}^{N_s} \sum_{l=1}^{N} \frac{\partial F_l(\mathbf{v}_i) |\mathbf{r}_l - \mathbf{r}_{0i}|^2}{\partial v_{km}^*}, \tag{B.6}$$

where

$$\frac{\partial F_l(\mathbf{v}_i)\,|\,\mathbf{r}_l-\mathbf{r}_{0i}\,|^2}{\partial v_{km}^*}=\frac{\partial F_l(\mathbf{v}_i)}{\partial v_{km}^*}\,|\,\mathbf{r}_l-\mathbf{r}_{0i}\,|^2+\frac{\partial\,|\,\mathbf{r}_l-\mathbf{r}_{0i}\,|^2}{\partial v_{km}^*}F_l(\mathbf{v}_i), \tag{B.7}$$

$$\frac{\partial F_l(\mathbf{v}_i)}{\partial v_{km}^*}=\frac{\partial \mathbf{q}_l^H\mathbf{v}_i^H\mathbf{v}_i\mathbf{q}_l}{\partial v_{km}^*}=\frac{\partial\left(\sum_{r=1}^{N_s}q_{lr}^*v_{ri}^*\right)}{\partial v_{km}^*}\mathbf{v}_i\mathbf{q}_l=\delta_{mi}q_{lk}^*\mathbf{v}_i\mathbf{q}_l, \tag{B.8}$$

and

$$\frac{\partial\,|\,\mathbf{r}_l-\mathbf{r}_{0i}\,|^2}{\partial v_{km}^*}=\partial\left|\mathbf{r}_l-\frac{\sum_{l=1}^{N}F_l(\mathbf{v}_i)\mathbf{r}_l\Delta\Gamma(\mathbf{r}_l)}{Tr\{\mathbf{\Gamma\Phi UDV}^H\mathbf{e}_i\mathbf{e}_i^T\mathbf{VDU}^H\mathbf{\Phi}^H\}}\right|^2\Big/\partial v_{km}^*, \tag{B.9}$$

with

$$Tr\{\mathbf{\Gamma\Phi UDV}^H\mathbf{e}_i\mathbf{e}_i^T\mathbf{VDU}^H\mathbf{\Phi}^H\}=\sum_{l'=1}^{N}F_{l'}(\mathbf{v}_i)\Delta\Gamma(\mathbf{r}_{l'}). \tag{B.10}$$

Combining Eqs. (B.9) and (B.10), one has

$$\partial\frac{\sum_{l=1}^{N}F_l(\mathbf{v}_i)\mathbf{r}_l\Delta\Gamma(\mathbf{r}_l)}{Tr\{\mathbf{\Gamma\Phi UDV}^H\mathbf{e}_i\mathbf{e}_i^T\mathbf{VDU}^H\mathbf{\Phi}^H\}}\Big/\partial v_{km}^*=\partial\frac{\sum_{l=1}^{N}F_l(\mathbf{v}_i)\mathbf{r}_l}{\sum_{l'=1}^{N}F_{l'}(\mathbf{v}_i)}\Big/\partial v_{km}^*,$$

$$=\left[\left(\sum_{l=1}^{N}\frac{\partial F_l(\mathbf{v}_i)}{\partial v_{km}^*}\mathbf{r}_l\right)\left(\sum_{l'=1}^{N}F_{l'}(\mathbf{v}_i)\right)-\left(\sum_{l=1}^{N}F_l(\mathbf{v}_i)\mathbf{r}_l\right)\left(\sum_{l'=1}^{N}\frac{\partial F_{l'}(\mathbf{v}_i)}{\partial v_{km}^*}\right)\right]\Big/\left(\sum_{l'=1}^{N}F_{l'}(\mathbf{v}_i)\right)^2 \tag{B.11}$$

which is easily calculated with the help of Eq. (B.8). Combining Eqs. (B.8), (B.9) and (B.11), the (m, k)-th element of the Euclidean gradient $\mathbf{E}_\Sigma$ can be deduced. Finally, the full Euclidean gradient $\mathbf{E}_\Sigma$ is given in the matrix format

$$\mathbf{E}_\Sigma=\frac{\partial\Sigma(\mathbf{V})}{\partial\mathbf{V}^H}=\begin{bmatrix}\frac{\partial\Sigma(\mathbf{V})}{\partial v_{11}^*} & \frac{\partial\Sigma(\mathbf{V})}{\partial v_{12}^*} & \cdots & \frac{\partial\Sigma(\mathbf{V})}{\partial v_{1N_s}^*}\\ \frac{\partial\Sigma(\mathbf{V})}{\partial v_{21}^*} & \frac{\partial\Sigma(\mathbf{V})}{\partial v_{22}^*} & \cdots & \frac{\partial\Sigma(\mathbf{V})}{\partial v_{2N_s}^*}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial\Sigma(\mathbf{V})}{\partial v_{N_s1}^*} & \frac{\partial\Sigma(\mathbf{V})}{\partial v_{N_s2}^*} & \cdots & \frac{\partial\Sigma(\mathbf{V})}{\partial v_{N_sN_s}^*}\end{bmatrix}. \tag{B.12}$$

The calculated Euclidean gradient $\mathbf{E}_\Sigma$ and the cost function $\Sigma(\mathbf{V})$ are embedded together into steps 2) and 7) of the CG algorithm introduced in subsection 3.3.1, and then the optimal matrix $\mathbf{V}$ is found by iterations. Accordingly, the mixing matrix $\mathbf{A}$ and the latent sources $\boldsymbol{\alpha}$ are respectively estimated with the help of Eqs. (2.10) and (2.19), and the optimal $\mathbf{V}$. At last, all separated sources are reconstructed based on the estimated $\mathbf{A}$ and $\boldsymbol{\alpha}$.

Appendix C

Optimization of CLSE in Stiefiel manifold

Similar to Eq. (B.1), the Euclidean gradient $\mathbf{E}_H$ of the cost function from the criterion of least spatial entropy is written as

$$\mathbf{E}_H = \frac{\partial H(\mathbf{V})}{\partial \mathbf{V}^H}. \tag{C.1}$$

First of all, substituting Eqs. (B.2), (B.3) and (B.10) into Eq. (3.29), the total spatial entropy $H(\mathbf{V})$ can be rewritten more concisely as

$$H(\mathbf{V}) = -\sum_{i=1}^{N_s}\sum_{l=1}^{N} F_l(\mathbf{v}_i)\Delta\Gamma(\mathbf{r}_l)\ln\left(\frac{F_l(\mathbf{v}_i)\Delta\Gamma(\mathbf{r}_l)}{\sum_{l'=1}^{N} F_{l'}(\mathbf{v}_i)\Delta\Gamma(\mathbf{r}_{l'})}\right). \tag{C.2}$$

The (m, k)-th element of the Euclidean gradient $\mathbf{E}_H$ can be formulated as

$$\begin{aligned}(\mathbf{E}_H)_{mk} &= \frac{\partial H(\mathbf{V})}{\partial v_{km}^*} \\ &= -\sum_{i=1}^{N_s}\sum_{l=1}^{N}\frac{\partial F_l(\mathbf{v}_i)}{\partial v_{km}^*}\left(1+\ln\left(\frac{F_l(\mathbf{v}_i)\Delta\Gamma(\mathbf{r}_l)}{\sum_{l'=1}^{N} F_{l'}(\mathbf{v}_i)\Delta\Gamma(\mathbf{r}_{l'})}\right)\right)\Delta\Gamma(\mathbf{r}_l). \\ &\quad - F_l(\mathbf{v}_i)\Delta\Gamma(\mathbf{r}_l)\frac{\sum_{l'=1}^{N}\frac{\partial F_{l'}(\mathbf{v}_i)}{\partial v_{km}^*}\Delta\Gamma(\mathbf{r}_{l'})}{\sum_{l'=1}^{N} F_{l'}(\mathbf{v}_i)\Delta\Gamma(\mathbf{r}_{l'})}\end{aligned} \tag{C.3}$$

With the help of Eq. (B.8), $(\mathbf{E}_H)_{mk}$ can be calculated. In the end, the full Euclidean gradient $\mathbf{E}_H$ is depicted as

$$\mathbf{E}_H = \frac{\partial H(\mathbf{V})}{\partial \mathbf{V}^H} = \begin{bmatrix} \frac{\partial H(\mathbf{V})}{\partial v_{11}^*} & \frac{\partial H(\mathbf{V})}{\partial v_{12}^*} & \cdots & \frac{\partial H(\mathbf{V})}{\partial v_{1N_s}^*} \\ \frac{\partial H(\mathbf{V})}{\partial v_{21}^*} & \frac{\partial H(\mathbf{V})}{\partial v_{22}^*} & \cdots & \frac{\partial H(\mathbf{V})}{\partial v_{2N_s}^*} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial H(\mathbf{V})}{\partial v_{N_s1}^*} & \frac{\partial H(\mathbf{V})}{\partial v_{N_s2}^*} & \cdots & \frac{\partial H(\mathbf{V})}{\partial v_{N_sN_s}^*} \end{bmatrix}. \quad \text{(C.4)}$$

In the same way as described at the end of Appendix A, the separated sources from the principle of least spatial entropy can be reconstructed with the help of the optimal $\mathbf{V}$.

Publications

Journals:

1. **Bin DONG**, Jérôme ANTONI, Erliang ZHANG, Blind separation of sound sources from the principle of least spatial entropy, Journal of Sound and Vibration 333 (9) (2014) 2643-2668.

2. Erliang ZHANG, Jérôme ANTONI, **Bin DONG**, Hichem SNOUSSI, Bayesian space-frequency separation of wide-band sound sources by a hierarchical approach, Journal of Acoustic Society of America 132 (5) (2012) 3240–3250.

Conferences:

1. **Bin DONG**, Jérôme ANTONI, Blind separation of uncorrelated sound sources from the principle of least spatial complexity, the 19th international congress on sound and vibration, Vilnius, Lithuania, July 08-12, 2012.

2. **Bin DONG**, Jérôme ANTONI, Blind separation of acoustic sources from the principle of least spatial complexity, the international conference of surveillance 6, Compiègne, France, October 25-26, 2011.

FOLIO ADMINISTRATIF

THESE SOUTENUE DEVANT L'INSTITUT NATIONAL DES SCIENCES APPLIQUEES DE LYON

NOM : **DONG** DATE de SOUTENANCE : 14 / 04 / 2014
(avec précision du nom de jeune fille, le cas échéant)

Prénoms : **Bin**

TITRE : Spatial Separation of Sound Sources

NATURE : Doctorat Numéro d'ordre :

Ecole doctorale : MEGA

Spécialité :Acoustique

RESUME :
Blind source separation is a promising technique for the identification, localization, and ranking of sound sources. The aim of this dissertation is to offer methods for separating incoherent sound sources which may overlap in both the space and frequency domains by exploiting spatial information. This is found of interest in acoustical applications involving the identification and ranking of sound sources stemming from different physical origins. The fundamental principle of all proposed methods proceeds in two steps, the first one being reminiscent to source reconstruction (e.g. as in near-field acoustical holography) and the second one to blind source separation. Specifically, the source mixture is first expanded into a linear combination of spatial basis functions whose coefficients are set by backpropagating the pressures measured by an array of microphones to the source domain. This leads to a formulation similar, but no identical, to blind source separation. In the second step, these coefficients are blindly separated into uncorrelated latent variables, assigned to incoherent “virtual sources”. These are shown to be defined up to an arbitrary rotation. A unique set of sound sources is finally recovered by searching for that rotation (conjugate gradient descent in the Stiefel manifold of unitary matrices) which minimizes some spatial criteria, such as spatial variance, spatial entropy, or spatial orthogonality. This results in the proposal of three separation criteria coined “least spatial variance”, “least spatial entropy”, and “spatial decorrelation”, respectively. Meanwhile, the condition under which classical decorrelation (principal component analysis) can solve the problem is deduced in a rigorous way. T he same concept of spatial entropy, which is central to the dissertation, is also exploited in defining a new criterion, the entropic L-curve, dedicated to determining the number of active sound sources on the source domain of interest. The idea consists in considering the number of sources that achieves the best compromise between a low spatial entropy (as expected from compact sources) and a low statistical entropy (as expected from a low residual error). The proposed methodology is validated on both laboratory experiments and numerical data, and illustrated on an industrial example concerned with the ranking of sound sources on the topside of a Diesel engine. The methodology can also correctly separate very small sources whose amplitudes are 40 dB lower than the strongest sources. At the same time, the robustness to the estimated number of active sources, to the range distance between the source domain of interest and the array of microphone, and to the size of aperture function is demonstrated with success.

MOTS-CLES : **inverse problem, backpropagation, blind source separation, principle of least spatial entropy, principle of least spatial variance, spatial decorrelation, joint approximate diagonalization**

Laboratoire (s) de recherche : Laboratoire Vibrations Acoustique

Directeur de thèse: Jérôme ANTONI

Président de jury : Charles Pezerat

Composition du jury :

Mohamed EL BADAOUI (Professeur), Université Jean Monnet - IUT de Roanne, Rapporteur
Jean-Hugh THOMAS (Maître de Conférences HDR), LAUM, Rapporteur
Charles PEZERAT (Professeur), LAUM, Examinateur
Quentin LECLERE (Maître de Conférences HDR), INSA de Lyon, Examinateur
Jérôme ANTONI (Professeur), INSA de Lyon, Directeur de thèse