On the representation of wavefronts localized in space-time and wavenumber-frequency domains

Cite as: JASA Express Lett. 1, 054801 (2021); https://doi.org/10.1121/10.0004852
Submitted: 24 November 2020. Accepted: 10 March 2021. Published Online: 05 May 2021

Elias Zea, and Marco Laudato

ARTICLES YOU MAY BE INTERESTED IN

Effects of Pacific Summer Water layer variations and ice cover on Beaufort Sea underwater sound ducting
The Journal of the Acoustical Society of America 149, 2117 (2021); https://doi.org/10.1121/10.0003929

Seminal article about model-based space-time array processing
The Journal of the Acoustical Society of America 149, R9 (2021); https://doi.org/10.1121/10.0004816

Machine learning in acoustics: Theory and applications
The Journal of the Acoustical Society of America 146, 3590 (2019); https://doi.org/10.1121/1.5133944

Across Acoustics
The official podcast highlighting authors’ research from our publications
On the representation of wavefronts localized in space-time and wavenumber-frequency domains

Elias Zea1,a) and Marco Laudato2,b)
1KTH Royal Institute of Technology, Department of Engineering Mechanics, The Marcus Wallenberg Laboratory for Sound and Vibration Research, SE-10044 Stockholm, Sweden
2Università degli Studi dell’Aquila, Dipartimento di Ingegneria e Scienze dell’Informazione e Matematica, Via Vetoio, L’Aquila, 67100 Italy
zea@kth.se, marco.laudato@graduate.univaq.it

Abstract: This Letter reports evidence suggesting a representation system for transient waves with band limited spectra, referred to here as localized waves in the space-time and wavenumber-frequency domains. A theoretical analysis with a transient monopole shows that the wavenumber-frequency pressure spectrum is distributed over hyperbolic regions of propagating waves and evanescent waves. An experimental analysis is performed, applying dictionary learning to reverberant sound fields measured with a microphone array in three rooms. The learned components appear to be related by analytical transformations in the spectra, suggesting a partitioning characterized by hyperbolic dispersion curves and multiple directions and times of arrival. © 2021 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

[Editor: Matthew A. Dzieciuch] https://doi.org/10.1121/10.0004852

Received: 24 November 2020   Accepted: 10 March 2021   Published Online: 5 May 2021

1. Introduction

Acoustic waves emitted by a transient sound source interact with the surrounding environment, undergoing a variety of transformations due to interference, scattering, and absorption.1 In practice, these transformations change the spectrum of the original wave, often resulting in a bandpass filtering effect in terms of incidence angles that are attenuated. An example is edge diffraction: transient waves scattered from a finite edge are confined to a localized region of space-time.2 If, additionally, the edge has a frequency-dependent impedance, even some wavelengths can be attenuated. In this context, we are interested in all such phenomena that result in transient acoustic waves that have a band limited spectrum: we shall refer to them as waves localized in space-time and wavenumber-frequency domains.

The notion of localization is present in the acoustics literature. One of the simplest examples is the case of harmonic plane waves—localized in the wavenumber-frequency domain, which are well suited to represent steady-state responses3 but ineffective to represent waves localized in space-time. Another example of space-wavenumber localization is the effect due to finite aperture and sensor density introduced by microphone array techniques such as near-field acoustic holography.4 Due to measurement constraints, the finiteness (density) of the array compromises the localization of waves in the wavenumber (space) domain.

The state-of-the-art of representation systems can be summarized as follows. To represent non-stationary acoustic waves, space-time can be partitioned into spatiotemporal patches, usually followed by localized Fourier transforms.5–7 A suitable choice of patch size allows one to localize the waves in space-time. Transient waves can also be represented in space and wavenumber domains with curvelets,6 without the possibility of modelling, however, band-limitedness in the frequency domain. As far as the authors know, the closest approach to representing acoustic waves localized in space-time and wavenumber-frequency domains is the use of shearlets to represent room impulse responses.5 Machine learning can also be used to train representations from acoustic data.9 Notable examples of application include the learning of underwater sound speed profiles using K-SVD,10 the use of sparse matrix factorization for scene classification,11 the estimation of scattering effects for spatial audio rendering,12 as well as the representation of steady-state sound fields in rooms with dictionary learning.13 Despite the fact that some of the state-of-the-art methods are able to represent (or learn to represent) acoustic waves, it is not clear how to analytically design a representation system departing from the underlying architecture of the data.

The main contribution of this Letter is to report the architecture of pressure data in the wavenumber-frequency domain, as a first step towards the analytical design of a representation system for localized waves. The hypothesis is that...

a)Author to whom correspondence should be addressed, ORCID: 0000-0001-5723-9571.
b)ORCID: 0000-0003-2153-9630.
the patterns governing this architecture follow a spectral partitioning into band limited regions of hyperbolic dispersion curves and intervals of phase velocities. The hypothesis is tested with a theoretical analysis of the pressure spectra of a transient monopole source, as well as with an experimental analysis applying dictionary learning on reverberant sound fields.

2. Theoretical analysis

Let us consider the dispersion relation of acoustic waves in a homogeneous medium at constant temperature and in the absence of flow

\[ (\omega/c)^2 = k_x^2 + k_y^2 + k_z^2, \]

(1)

which is obtained by Fourier transforming the linearized wave equation in Cartesian coordinates: \( \partial_t P = c^2 \nabla^2 P \), where \( P \) is the sound pressure and \( t \) is time. Here, the constant \( k = \omega/c \) is the acoustic wavenumber, \( \omega \) is the angular frequency, \( c \) is the sound speed, and \( k_i \) is the wavenumber along the \( i = x, y, z \) spatial coordinates. Assuming without loss of generality that waves propagate along the \( z \)-axis, their wavenumber is \( k_z = \pm \sqrt{k^2 - k_x^2 - k_y^2} \). This leads to the concept of radiation circle: \( k^2 = k_x^2 + k_y^2 \). Acoustic waves are said to propagate to the far-field when \( k_z \) is real, i.e., \( k^2 \geq k_x^2 + k_y^2 \), while evanescent waves decay exponentially in the near-field when \( k_z \) is imaginary, i.e., \( k^2 < k_x^2 + k_y^2 \). By letting \( k \) vary, the radiation circle becomes a radiation cone; in other words, the radiation cone is defined as the extrusion of the radiation circle with a variable frequency \( \omega \in (-\infty, \infty) \).

2.1 Analysis with a transient monopole

Let us consider two spatial dimensions \( (x, y) \) and time \( t \). Assume a monopole sound source, located at \( (x_s, y_s) \), is emitting a transient pulse towards a continuous, infinite line of receivers located at \( y = 0 \), as shown in Fig. 1(a). The pulse radiated by the source is \( P(x, y, t) = \delta(t - r/c)/(4\pi r) \), where \( r = \sqrt{(x - x_s)^2 + (y - y_s)^2} \). Translations of the line of receivers along the \( x \) and \( y \) axes correspond to trivial (phase) transformations in the wavenumber-frequency domain \( (k_x, \omega) \). A non-trivial transformation in space-time that leads to a modulus change in the wavenumber-frequency domain is a rotation around the normal to the \( (x, y) \) plane [see Fig. 1(b)].

Consider a rigid rotation with an angle \( \alpha \in (-\pi/2, \pi/2) \). The distance between the source and a generic point \( (x \cos \alpha, x \sin \alpha) \) on the line of receivers is now a function of \( x \),

\[ r(x, z) = \left\{ \left[ (x - (x_s \cos \alpha + y_s \sin \alpha))^2 + x_s^2 + y_s^2 - (x_s \cos \alpha + y_s \sin \alpha)^2 \right]^{1/2} \right\}. \]

(2)

The Fourier transform of the pressure with respect to space \( x \) and time \( t \), as a function of \( x \), reads

\[ \tilde{P}(k_x, \omega; z) = \frac{1}{4\pi^2} \delta \left( t - \frac{r(x, z)}{c} \right) e^{-ik_x + i\omega t} \frac{dx dt}{r(x, z)} \]

(3)

By defining \( r_1(x) = x \cos \alpha + y \sin \alpha \), \( u = x - r_1(x) \), and \( g^2(x) = u^2 + g^2(x) \), then \( r(x, z) = \sqrt{u^2 + g^2(x)} \), and the integral becomes

\[ \int_{-\infty}^{\infty} \frac{e^{-ik_x + i\omega t}}{r(x, z)} \frac{dx dt}{r(x, z)}. \]

Fig. 1. Geometrical setup with monopole source (orange circle) and continuous receiver line (solid black line). (a) Receiver line at \( y = 0 \). (b) Receiver line tilted with respect to the \( x \) axis by \( \alpha \).
\[
\tilde{P}(k_x, \omega; z) = \frac{e^{-jk_xv_0(z)}}{2\pi} \int_{0}^{\infty} \frac{e^{-j(\omega/c)s} \sqrt{u^2 + g^2(z)}}{\sqrt{u^2 + g^2(z)}^2} \cos(k_xu) du.
\] (4)

From the table of integrals, the pressure spectrum becomes

\[
\tilde{P}(k_x, \omega; z) = -\frac{j \epsilon^{-jk_xv_0(z)}}{4} H_0^{(1)}(g(z) \sqrt{\frac{(\omega/c)^2 - k_x^2}{\omega/c}}).
\] (5)

where \(H_0^{(1)}\) is the complex conjugate of the zero-th order Hankel function of the first kind. The pressure spectrum is singular whenever the argument of \(H_0^{(1)}\) is zero: and this happens whenever Eq. (1) is satisfied, i.e., at the radiation cone \((\omega/c)^2 = k_x^2\) and/or when \(g(z) = 0\) where \(z = \tan^{-1}(y/x)\).

2.2 Partitioning the wavenumber-frequency domain

Let us now analyze Eq. (5) more into detail by considering four values of \(z\). Figures 2(a)–2(d) show the corresponding sound pressure level spectra. It can be observed that there are contours of constant sound pressure level, which follow hyperbolic trajectories asymptotically approaching the radiation cone (depicted in diagonals). One can also observe that the spectrum becomes increasingly broadband as \(z\) approaches \(\pi/2\) or, more generally, \(z\). Figures 2(e)–2(g) show the partitioning of the wavenumber-frequency domain of three state-of-the-art representation systems: (e) Cartesian, (f) directional sub-bands, and (g) shearlets. It can be seen that these partitioning approaches do not follow closely the structure of the spectra in Figs. 2(a)–2(d).

We propose a different partitioning of the wavenumber-frequency domain, shown in Fig. 2(h), into band limited regions described by hyperbolic dispersion curves and lines of constant phase velocity. The hyperbolic curves are naturally linked with the physics of Eq. (5), and can be described analytically by bounding the square root in the argument of the Hankel function to intervals \([a, b]\), \(0 < a < b\); which gives the hyperbolic regions \(a \leq |(\omega/c)^2 - k_x^2| \leq b\). In order to make these hyperbolic regions band limited, i.e., able to represent band limited waves, we need lines of constant phase velocity. The physical motivation is that sound sources have directivity patterns in practice, resulting in band limited spectra within intervals of phase velocities. This is equivalent to the so-called directional sub-bands, shown in Fig. 2(f), which are closely related to the direction and time of arrival of the waves. These intervals of phase velocities can be described analytically as follows: \(\omega/v_1 \leq k_x \leq \omega/v_2\), \(v_1 > v_2 > 0\).

3. Experimental analysis

The experimental data consists of three sets of room impulse responses measured with a 1D microphone array. Reverberant sound fields comprise rich training data involving absorption, diffraction, and interference phenomena, which results in waves with various wavelengths, directions and times of arrival. An omnidirectional source is placed in three furnished rooms, at 1.5 m height and 2 m away from the array. The spatiotemporal impulse responses are obtained with sine sweeps in the frequency bandwidth 100–4500 Hz, with sampling rate \(f_s = 11.25\) kHz. The array is placed at 1.2 m height, and has 100 positions spaced by \(\Delta x = 3\) cm, resulting in a wavenumber bandwidth of 209 rad/m.

![Fig. 2. Top: Wavenumber-frequency spectra of sound pressure level in dB re 20 μPa, via Eq. (5) with \(c = 1\) m/s, \((x_s, y_s) = (0, 1)\) m, for (a) \(x = 0\) rad, (b) \(x = \pi/4\) rad, (c) \(x = \pi/3\) rad, and (d) \(x = 0.85\pi/2\) rad. Bottom: illustration of state-of-the-art partitioning approaches, (e) Cartesian, (f) directional sub-bands (Ref. 6), (g) shearlets (Ref. 9), and (h) proposed partitioning.](Image)
The space-time datasets are first pre-processed for dictionary learning. The data is divided into non-overlapping patches of $32 \times 32$ space-time samples each. A Tukey window and a 2D Fourier transform is applied to each patch, resulting in pressure spectra patches in the wavenumber-frequency domain. The wavenumber-frequency resolution of each patch is $\Delta k = 6.5 \text{ rad/m}$ and $\Delta \omega = 2209 \text{ rad/s}$. Each spectrum is then vectorized and allocated into the columns of an observation matrix $\mathbf{B} \in \mathbb{C}^{M \times N}$, where $M = 1024$ is the number of space-time samples, and $N = 1485$ is the total number of space-time patches (each of size $32 \times 32$). All impulse responses, approximately $1.5 \times 10^6$ samples, are considered for the training set.

The pre-processed data is then fed to the dictionary learning algorithm. The main idea is to learn a limited number of components, defined over wavenumber-frequency patches of size $32 \times 32$, that best represent the variance of the data. There is a variety of dictionary learning algorithms, and in this Letter we choose sparse principal component analysis. In contrast to conventional principal component analysis (PCA), sparse PCA learns a redundant set of loading vectors, i.e., no longer orthogonal, and it ensures that each data point can be represented using a small number of components. Redundant representations are known to benefit from better performance against noise and missing data, as well as a more parsimonious representation of the data.

The dictionary learning is formulated as the following optimization problem:

$$
\min_{\mathbf{U}, \mathbf{w}} \sum_{i=1}^{N} \frac{1}{2} | | \mathbf{b}_i - \mathbf{U} \mathbf{w}_i | |_2^2 + \mu | | \mathbf{w}_i | |_1^p,
$$

which is solved by means of a soft thresholding algorithm using the MATLAB toolbox SPASM. Here, $\mathbf{b}_i \in \mathbb{C}^M$ is the $i$th column of the observation matrix $\mathbf{B}$, $\mathbf{U} \in \mathbb{C}^{M \times Q}$ is the dictionary matrix whose columns are the $Q$ learned components, $\mathbf{w}_i \in \mathbb{C}^Q$ contains the weights of the components, $\mu$ is a regularization parameter, and $| | \cdot | |_p = (\sum | | \cdot | |_p^q)^{1/p}$ is the vector $\ell_p$ norm, $0 < p < \infty$. The choice of regularization parameter is $0 < \mu < M$, and it determines the sparsity of the learned components: $\mu$ out of $M$ non-zero coefficients.

Figures 3(a)–3(r) show the magnitude of $Q = 18$ learned components, with 12.5% non-zero coefficients; while Fig. 3(s) shows the variance of the learned components versus $Q$, with 12.5% and 50% non-zero coefficients. It can be seen that a less sparse (denser representation) requires more learned components (higher $Q$) in order to explain the same variance. Moreover, the decay of the variance is irregular for $Q < 15$, and it becomes nearly exponential for $Q \geq 15$.

The components plotted in Figs. 3(a)–3(r) have overlaid the cone, and the proposed partitioning shown in Fig. 2(h). The magnitude of the learned components is symmetric [mirrored with respect to the $(k_x, \omega)$ axes], attributed to the

---

Fig. 3. (a)–(r) Normalized magnitude of the wavenumber-frequency spectra of $Q = 18$ sparse principal components, with 12.5% non-zero coefficients. Overlaid: cone, and proposed partitioning [Fig. 2(h)]. The axes ($k_x, \omega$) are normalized, and the colormaps are interpolated. Boxes A, B, C, and D indicate specific regions in the spectra. (s) Variance of the learned components versus $Q$, with (solid) 12.5% and (dotted) 50% non-zero coefficients. The dashed vertical line indicates the variance for $Q = 18$ components. The variance is an adjusted variance that accounts for the correlation across components.
fact that the training data is real. Moreover, components are learned both in the far-field and the near-field regions of the wavenumber-frequency domain. In particular, the evanescent components learned are localized in low frequency regions, which is physical since the measurement array is not too close to the boundaries of the room, and near-fields decay stronger with increasing frequency.

The hypothesis under study can be examined with the results in Figs. 3(a)–3(r). To facilitate the analysis, we have highlighted band limited regions in the spectra with boxes A, B, C, and D. How close the learned components follow the proposed partitioning can be studied by looking at how the components relate to each other. For instance, the regions A, B, and C, appear to be related by means of a curved transformation in \((k_x, \omega)\) from one component to another [e.g., compare Fig. 3(b) with Fig. 3(e) and Fig. 3(b) with Fig. 3(o)]. This indicates a strong structure is present in the data, and suggests a resemblance to the hyperbolic dispersion curves observed in Figs. 2(a)–2(d). A similar observation can be drawn for the near-field regions indicated by D in Figs. 3(i) and 3(p). Most of the remaining components [see, e.g., Figs. 3(i) and 3(ii) and Figs. 3(g) and 3(m)], also contain regions related by such curved transformations. More evidently, the components are also shifted diagonally within band limited intervals of phase velocities. These intervals correspond to multiple wave directions and times of arrival in the array, which is a natural feature of the training data: reverberant sound fields in rooms.

The outcome of the dictionary learning yields some components that have combinations of active regions in the partitioning: we refer to these as “interfering components.” For example, in Fig. 3(p), two distinct band limited components can be seen in the right-hand side of the spectrum (one of which is indicated by region D), which have their symmetric counterparts with respect to \((k_x, \omega)\) in the left-hand side of the spectrum. This suggests the interference of band limited waves with two distinct directions of arrival. Analogously in Fig. 3(b), two band limited components (A and B, plus their symmetric counterparts) are seen within the same phase velocity interval, which suggests the interference of waves with the same direction of arrival but two distinct wavelengths. It should be noted, however, that a representation system including interfering components would be ineffective at representing individual waves. Therefore, future work handcrafting the representation system should involve only individual waves, the linear superposition of which results in their interference.

4. Conclusions

This Letter shows evidence suggesting a new representation system for transient band limited acoustic waves. A theoretical analysis of the transient pressure emitted by a point source shows that the wavenumber-frequency spectrum is distributed along regions of hyperbolic dispersion curves. An experimental analysis is performed using dictionary learning of reverberant sound fields measured with a microphone array. These analyses suggest a new spectral partitioning into band limited regions described by hyperbolic dispersion curves and intervals of phase velocities. The promising result is that the components learned in the experimental analysis appear to be related by analytical transformations of these band limited regions in the spectra. Future work involves handcrafting a representation system based on such analytical transformations, which we suspect might lead to an efficient implementation and a sparser representation than the state-of-the-art approaches.

Acknowledgments

This work has been financially supported by Vetenskapsrådet under Grant Agreement No. 2020–04668. We want to thank the reviewers for the valuable comments and suggestions. We are grateful for the discussions and feedback from Efren Fernandez-Grande, Joakim Andén, Ines Lopez Arteaga, Özän Öktem, Can Evren Yarman, Romain Rumpler, and Ciarán O’Reilly.

References and Links

1H. Kuttruff, *Room Acoustics*, 5th ed. (Spon Press, London, 2009).
2V. Pulkkki and U. P. Svensson, “Machine-learning-based estimation and rendering of scattering in virtual reality,” *J. Acoust. Soc. Am.* 145(4), 2664–2676 (2019).
3E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic Press, San Diego, 1999).
4J. D. Maynard, E. G. Williams, and Y. Lee, “Nearfield acoustic holography: I. Theory of generalized holography and the development of NAH,” *J. Acoust. Soc. Am.* 78(4), 1395–1413 (1985).
5J.-H. Thomas, V. Gruler, S. Paillasseur, J.-C. Pascal, and J.-C. Le Roux, “Real-time near-field acoustic holography for continuously visualizing nonstationary acoustic fields,” *J. Acoust. Soc. Am.* 128(6), 3554–3567 (2010).
6F. Pinto and M. Vetterli, “Space-time-frequency processing of acoustic wave fields: Theory, algorithms, and applications,” *IEEE Trans. Sign. Process.* 58(9), 4608–4620 (2010).
7M. Nolan, M. Berzborn, and E. Fernandez-Grande, “Isotropy in decaying reverberant sound fields,” *J. Acoust. Soc. Am.* 148(2), 1077–1088 (2020).
8E. J. Candes and L. Demanet, “The curlvector representation of wave propagators is optimally sparse,” *Commun. Pure Appl. Anal.* 58(11), 1472–1528 (2005).
9E. Zea, “Compressed sensing of impulse responses in rooms of unknown properties and contents,” *J. Sound Vib.* 459, 114871 (2019).
10M. J. Bianco, P. Gerstoft, J. Traer, E. Oznich, M. A. Roch, S. Gannot, and C.-A. Deledalle, “Machine learning in acoustics: Theory and applications,” *J. Acoust. Soc. Am.* 146(5), 3590–3628 (2019).
11M. Bianco and P. Gerstoft, “Dictionary learning of sound speed profiles,” *J. Acoust. Soc. Am.* 141(3), 1749–1758 (2017).
12. V. Bisot, R. Serizel, S. Essid, and G. Richard, “Feature learning with matrix factorization applied to acoustic scene classification,” IEEE/ACM Trans. Audio Speech 25(6), 1216–1229 (2017).

13. M. Hahmann, S. Verburg, and E. Fernandez-Grande, “Analysis of a sound field in a room using dictionary learning,” in Proceedings of the 23rd International Congress on Acoustics (Deutsche Gesellschaft für Akustik, Berlin, 2019). pp. 149–154.

14. I. Gradshteyn and I. Ryzhik, Table of Integrals, Series, and Products, 4th ed. (Academic, New York, 1965).

15. M. Elad, Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing (Springer, London, 2010).

16. P. G. Casazza and G. Kutyniok, Finite Frames: Theory and Applications (Springer Science+Business Media, London, 2013).

17. K. Sjöstrand, L. Clemmensen, R. Larsen, G. Einarsson, and B. Ersbøll, “SpaSM: A MATLAB toolbox for sparse statistical modeling,” J. Stat. Softw. 84(10), 1–37 (2018).