Bibliographical reference

Reconstructing Stimulus Space Geometry from Auditory Cortex Responses, Bernhard Englitz and Shihab Shamma, Acta Acustica **vol. 104** (Number 5), 2018, pp. 865-869

DOI

https://doi.org/10.3813/AAA.919234

Acta Acustica united with Acustica
Vol. 104 (2018) 865 – 869

DOI 10.3813/AAA.919234

Reconstructing Stimulus Space Geometry from Auditory Cortex Responses

Bernhard Englitz[1)], Shihab Shamma[2)]
[1)] Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen,
The Netherlands. b.englitz@donders.ru.nl
[2)] Ecole Normale Superieure, Paris, France; University of Maryland, College Park, USA

Summary

When comparing different stimuli one often needs to estimate distances between them, for example differences in frequency of tones. If one considers the distances between all stimuli, then the stimuli can be organized in a geometric arrangement. For example, pure tones are defined by their frequency and thus form a straight line, while Shepard tones are identical after an octave shift and thus form a circle. Previously, representational similarity analysis (RSA) was proposed to infer the geometric relations between neurally encoded stimuli from distances between their population responses. Applied to single unit responses of ferret auditory cortex, we find that RSA produces a warped large-scale geometry of the stimulus set. We find this to be caused by spectrally localized neural responses, as is typical in primary auditory cortex. We extend RSA to global RSA (gRSA) by estimating global distances from local distances, using stochastic integration along stimulus neighborhoods. We demonstrate that gRSA can distinguish the physical, overall geometries for pure and Shepard tones. The proposed approach is relevant in many auditory tasks such as estimating sound direction (2D sphere) and position (3D space) or segregating streams in auditory scenes.

PACS no. 43.66.Ba, 43.66.Bf, 43.66.Ki, 43.66.Lj

1. Introduction

In the process of neural encoding, external stimuli are mapped to internal, neural representations. Many real world tasks involve comparing stimuli, e.g. to assess if two stimuli should be grouped together (*classification*) or how different two stimuli are (*measurement*). To assess these relations between the external stimuli, the nervous system can only rely on its internal representations. Distances in the internal representation can be used as a proxy to measure distances between external stimuli. More generally, the geometical arrangement of a whole set of external stimuli can be characterized by the set of distances between its neural representations.

This conceptual approach to understanding both perceptual and neural representations - termed similarity analysis (see [1] for review) - was pioneered early on by Torgerson and Shepard [2, 3] via the development of *multidimensional scaling* (MDS), a general technique that produces a low-dimensional geometrical representation from a set of distances, which had considerable influence in psychophysics. Famous examples include the recreation of color and consonant perceptual spaces from human interstimulus distance estimates (for review see [4]).

More recently, Kriegeskorte [5] pioneered the application of similarity analysis to neural responses, in particular from the inferotemporal cortex to analyze the response to different visual objects in both humans and macaques [6]. This formulation of similarity analysis on the basis of neural representations was termed Representational Similarity Analysis (RSA). It consists of estimating distances between population responses to different stimuli, and applying MDS to the resulting set of distances.

When using RSA to analyze neural responses from ferret primary auditory cortex, we find the resulting stimulus space estimates to correctly reflect nearby stimulus distances. However, it fails to match the distances between physically more distant stimuli, as the overall geometry of the stimulus set is not faithfully reconstructed. Concretely, very low and very high frequency tones are reconstructed surprisingly close to each other, while humans can estimate both small (e.g. two tones which are a semitone apart) and large (e.g. tones which are multiple octaves apart) differences in frequency. We trace the source of this problem to the response properties of primary auditory cortex neurons, which mostly cover a limited range of pure tone frequencies. Consequently, mostly interstimulus distances of nearby stimuli can be measured accurately.

Received 5 May 2018,
accepted 15 August 2018.

Devising a novel stochastic integration technique between population responses of individual stimuli, we recover the large-scale distances between all stimuli. We apply this generalized, global RSA (gRSA) to pure tones and Shepard tones, correctly recovering the corresponding physical stimulus geometries, i.e. linear and circular, respectively. These results underscore the generality of similarity analysis for the analysis of basic and complex stimulus classes in audition and other senses, and suggest an important extension of RSA in order to cope with local stimulus responses. Relevant application domains include the estimation of relative object distances in space as well as parsing an auditory scene with multiple concurrent streams.

2. Methods

The animal experiments were conducted according to the ethical guide-lines of the University of Maryland and approved by the local IACUC.

2.1. Acoustic Stimuli and Presentation

Stimuli were presented in a sound-attenuated chamber (mac3, Industrial Acoustics Corporation). All stimuli were composed of randomized sequences of pure and Shepard tones [7]. A Shepard tone is a complex tone built as the sum of octave-spaced pure tones (using a flat spectral envelope). A Shepard tone can be characterized by its position in an octave, termed pitch class (in units of semitones), w.r.t. a base tone. Across the entire set of experiments the duration of the tones was 0.1 s at an overall level of 70 dBSPL, and each tone presented 5-20 times. The impulse response of the speaker was compensated to achieve a flat calibration over the range of 0.08-35 kHz by convolving the stimulus with the inverse of the speaker's impulse response. Stimuli were generated in Matlab using custom-written code sampled at 100 kHz, and then D/A converted via a DAQ-card (NI-6259, National Instruments). The analog signal was amplified using an analog amplifier (Crown D75A) and transduced via a speaker (Manger).

2.2. Neural Recordings

Extracellular neuronal recordings were collected from awake ferrets using multielectrode recordings sampled at 25 kHz from primary auditory cortex A1. While the recording array was always centered on A1, at the edges neurons from other areas of auditory cortex may have been included occasionally [8]. For details on the recording techniques see elsewhere [9]. For the present analysis, neural responses were collected across multiple recording sites and animals for matched stimuli to obtain a population response for each stimulus.

2.3. Data Analysis

First, recordings were bandpass (300-7000 Hz) filtered, and spike-sorted using custom-written code. Responses were then aligned to stimulus onsets and collected and averaged across stimulus repetitions. Collecting the responses for all stimuli one obtains the so-called encoding matrix (Figure 1 top). For visual convenience, the encoding matrix is here shown sorted by stimulus frequency. However, the analysis below does not take sorting into account.

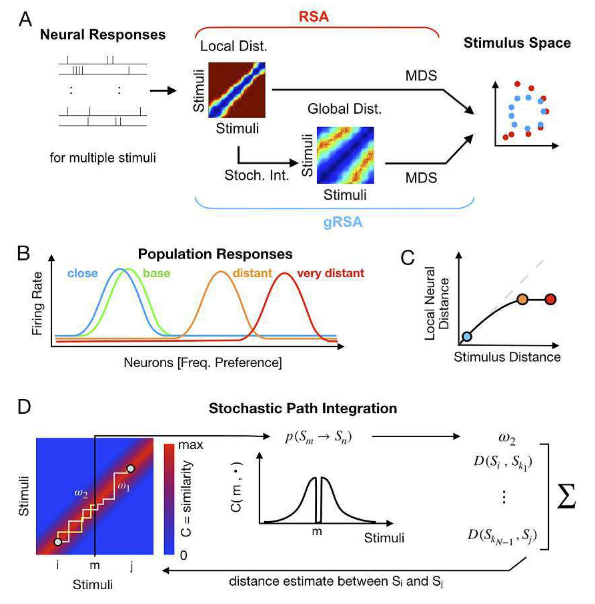

Figure 1. (A) Extending RSA to gRSA: While classical RSA performs MDS on the basis of local distances between stimuli (top), gRSA uses also large-scale distances estimated through stochastic integration in local neighborhoods (bottom, see D). (B) When comparing the distance between population response which do not span the entire population, distant and very distant stimuli are equally dissimilar from a given base stimulus. (C) Hence, large distances can be underestimated in comparison to the physical stimulus distance. (D) Illustration of the estimation of stimulus distances based on stochastic integration along local neighborhoods (see Methods for details).

In classical RSA (Figure 1A top), a distance matrix between stimuli is defined as

$$D(i, j) = 1 - C(i, j)^+ \tag{1}$$

where $C(i, j) := \sum_{n \in \text{Neurons}} \mathbf{N}_i(n)\mathbf{N}_j(n)$ is the correlation between population responses N_i and N_j for different stimuli S_i and S_j, respectively, and the '+' indicates taking only the positive part [6]. Hence, the distance between stimuli is measured as the degree of difference in their population responses. $D(i, j)$ is then processed with MDS, which provides an embedding of the stimuli in a lower dimensional space, largely retaining the distances [2, 4].

If stimuli only evoke responses in a restricted set of neurons, $D(i, j)$ contains only information about local dissimilarity, i.e. between nearby stimuli, since beyond a certain level of dissimilarity the entries of $D(i, j)$ are uniformly 0. Hence, distances between stimuli that are larger than the 'width' of the population response are not correctly represented. This already follows from one of the central theorems in MDS: in order to correctly recreate the Euclidian distances between a set of objects, one needs to have a complete rank-ordering of the interobject distances [10]. However, if population response are not broad enough, then one does not obtain a rank ordering for larger distances between stimuli (see Figure 1B).

To address this problem, we generalize RSA by estimating a global, large-scale distance matrix $G(i,j)$ on the basis of $D(i,j)$ using stochastic path integration along neighborhoods of stimuli (Figure 1C). More precisely, the values in $C(i,j)$ are interpreted as a probability density for considering stimulus j to be a direct neighbor of stimulus i, at distance $D(i,j)$. For each pair of stimuli (S_i, S_j), one starts at $S_i = S_{k_0}$ and follows a chain of probabilistic transitions to putative neighbors S_{k_1}, S_{k_2}, ... until $S_{k_n} = S_j$, where each transition is given by $p(S_m \rightarrow S_n) \propto C(i,j)^+$. To limit the number of possible paths, only correlations > 0.7 were considered. The large-scale distance estimate for a given run is then given as $d(i,j,\omega) = \sum_{l=1...n} D(k_{l-1}, k_l)$, where ω indicates the randomness of the steps. Performing this stochastic integration many times we can define $G(i,j) = \min_{\omega}(d(i,j,\omega))$. The path length estimation becomes more robust against noise in the neural activity, if instead of the minimum a low percentile (e.g. 5th) is used. Two paths ω_1 and ω_2 are illustrated in Figure 1D. They follow the local neighborhoods since steps are more likely close to the diagonal, formed by population responses of closeby stimuli. However, as mentioned above, if the stimuli were not sorted, the same paths would be chosen, but their display would look less intuitive. As in RSA, MDS is then performed in gRSA on $G(i,j)$ to obtain a low-dimensional embedding of the stimulus set. Since the distances in $G(i,j)$ are now better ordered, the Euclidean distances after MDS are also better preserved.

3. Results

We reanalyzed previously obtained recordings from the auditory cortex of awake ferrets. Responses were obtained for pure tone stimuli and Shepard tone stimuli. For pure tone stimuli we sampled from a total set of 3200 recordings, while for Shepard tones 919 recordings were available. Not all of these recordings contained units, and if a unit was contained it was often not tuned. Cells were selected for analysis if they exhibited a tuning to pure tones (only 100, to reduce computation time) or Shepard tones (42), significantly above their individual spontaneous firing rate. This could include both unimodal and multimodal tuning curves, although the vast majority of tuning curves were unimodal. The responses were arranged as a matrix of stimuli versus neurons, sorted by frequency. As detailed in the methods, gRSA does not utilize this sorting.

First, we analyze the population responses to simple pure tones. In classical RSA the distance matrix provides useful information mostly around the diagonal, i.e. for pure tones which are close in frequency (Figure 2A1). In gRSA, the distance matrix resulting from stochastic exploration (1000 random trajectories) shows an increase in distance away from the diagonal (Figure 2B1).

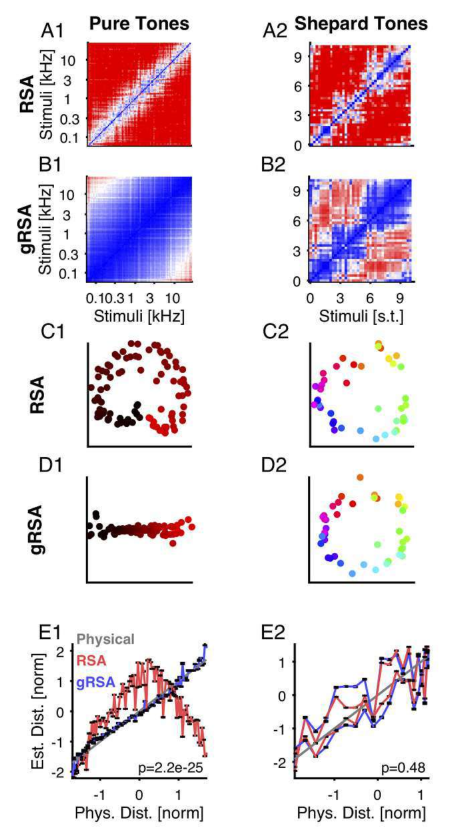

Figure 2. gRSA correctly reconstructs the stimulus geometry of pure and Shepard tones. **(A1)** The distance matrix for RSA is only locally informative, i.e. distant stimuli are estimated to the equally distant. **(B1)** In comparison the gRSA distance matrix already captures the linear increase in distance orthogonal to the diagonal. **(C1)** After MDS, the estimated geometry in RSA exhibits a globally curved nature, since large distances are not correctly represented (color indicates pure tone frequency from low (black) to high (red)). **D1** In gRSA's estimate the linear nature is well preserved. **(E1)** Distances estimated by RSA are much less accurate that those by gRSA (n=10 runs, 1000 stochastic trajectories, Wilcoxon rank sum test). **(A2-E2)** For Shepard tones, analogous analyses were performed. While the distance matrix used by RSA is not true to the actual distances (**A2**), the reconstruction still suceeds **(C2/E2)**, since a circle is the best approximation (although this will fail in general). gRSA arrives at a much more accurate estimation of the distances (**(B2)** sinusoidal increase and decrease orthogonal to the diagonal), and arrives at the same circular geometry **(D2)** . Note, that the estimates **(E2)** are quite matched in detail, which probably is based on particularities of set of neural responses.

Classical RSA reconstructs the geometry of the pure tone stimuli as a curved line in the surrounding space (Figure 2C1, color scheme: from low (black) to high (red) frequencies; plot axes are left unlabelled as these dimensions are not associated with physical units). While the local geometry is 1D in line with the single governing parameter (frequency), the global geometry is curved. As a consequence the Euclidean distances in 2D space differ markedly from the distances between the physical stimuli (Figure 2E1, red). Distance is here measured from the lowest frequency pure tone to all other tones (i.e. 80-32000 Hz). Applying MDS to the gRSA distance matrix can use the integrated distances between stimuli and hence arrives at a linear geometry also on the larger scale (Figure 2D1). Therefore the large-scale distances are correctly estimated, and the estimate coaligns with the physical distances. Since the scale of MDS is not coscaled with the actual frequencies, each estimate and the real distances were Z-scored. The difference in reconstruction quality of the distances between RSA (RMSE from diagonal = 0.92 ± 0.07) and gRSA (0.13 ± 0.01) is highly significant ($p < 10^{-24}$, Wilcoxon rank sum test, across all stimuli).

The insufficient separation of the ends of the frequency range is not a direct consequence of the audible frequency range: while the population responses go to their spontaneous values for inaudibly low and high tones, the correlation between these responses is low, translating to a large distance in the RSA's distance matrix (see Eq.1). However, this distance is comparable to distances between population responses between distant low and high frequency tones in the audible range.

Second, we analyse the neural responses to Shepard tones. Since a translation of a Shepard tones by one octave leads to the same physical stimulus, we expect the corresponding stimulus space to be circular, consistent with human distance estimation of distances between Shepard tone pairs (see e.g. [11]).

The local distances used by RSA are only consistent with a closed line, due to the wrapping at the ends of the range (top/bottom, Figure 2A2). Classical RSA therefore arrives at reconstructing a close line (Figure 2C2). The large-scale distance estimation in gRSA provides a good estimate of the stimulus-stimulus distances, leading to an approximately sinusoidal variation of distances, orthogonal to the diagonal of the distance matrix (Figure 2C2). Consequently, the overall arrangement of the Shepard tone stimulus space is reconstructed as a circle (Figure 2D2). Comparing the reestimated distances between RSA (0.43 ± 0.05) and gRSA (0.50 ± 0.06) against the distances in a regular circular arrangement shows overall good agreement, also between the methods (p=0.48, Wilcoxon rank sum test, as above, see Figure 2E2). The residual variations are probably due to variability in the neural responses, which would vanish for larger sample of neurons.

In summary, as predicted from its reliance on differences between population responses of similar stimuli, RSA failed to recover the large-scale structure for the linear stimulus set, warping it in the surrounding space. If measurements of distance are performed in this surrounding space, the results are highly inconsistent with the physical distances of the stimulus. In contrast, gRSA recovered the overall geometry correctly. Since the estimation of the large-scale distances is a stochastic process, the results of gRSA are a bit noisier for the Shepard tones, which could be remedied with a larger set of well-tuned neurons.

4. Discussion

Interpreting the internal neural representations in a way that represents meaningfully the relations of their external causes is essential for an effective neural system. We here presented an extension of RSA in order to correctly estimate larger distances and the overall geometry of stimulus spaces for the example of pure and Shepard tones.

For the reconstruction of the pure tones, we find a linear reconstruction on the basis of gRSA, and judged this reconstruction to be more faithful. For the case of the ferret primary cortical area, with largely monotonically tuned neurons, a linear geometry appears indeed most appropriate, as different stimuli will be further and further separated from each other, not only physically in their defining frequency, but also physically in a cortical area based on the tonotopic arrangement (compare e.g. [8]). However, this will not generally reflect the situation in other cortical areas, e.g. when neurons span multiple octaves, or the stimulus space reconstruction one would obtain from a human, well acquainted with musical harmonies. In fact, an interesting human study would be to record from multiple cortical areas (using fMRI), and reconstruct the stimulus space from each one separately, to detail the transformation from the physical to the psychophysical representation. The latter would be expected to resemble a helix, i.e. with overall increasing distances, but relatively shorted distances at the octaves.

For both RSA and gRSA an important question is the potential neural implementation of MDS. While it is currently unknown whether MDS is neurally realized, multiple groups have proposed neural implementations, e.g. [12]. For gRSA a stochastic integration step is introduced here. The most obvious way of implementing it would be to perform one stochastic estimation for each presentation of a pair of stimuli, to incrementally update the final distance matrix $G(i, j)$. The latter could be realized as the weight vectors of a global distance readout of the primary sensory cortices, realized also as a population code, which would only scale linearly in the represented stimuli.

The results of both RSA and gRSA can be quite sensitive to neurons with complex response properties. While we focussed here on primary response types, i.e. with typically monomodal tuning curves, higher auditory regions may have multi-peaked tuning curves ([13]). Such neurons create 'shortcuts' between different stimuli, since they correlate well with other neurons at both of their peak locations. Whether distance estimation in the brain is restricted to neurons with simple response properties or whether these 'shortcuts' correspond to actually perceived similarities between stimuli is an important question that could be investigated in human/animal psychophysics.

While the current analysis focussed on a simple set of acoustic stimuli that exemplify the differences between RSA and gRSA in a minimalistic fashion, similarity analysis could be used to reinterpret other classical topics in hearing, in particular stream segregation and spatial hearing.

Stream segregation is a minimal version of auditory scene analysis [14], i.e. the task of decomposing a complex acoustic environment into its contributing sources. This task can be conceptualized from the perspective of similarity analysis as identifying sources as parts of the stimulus which are distant from each other. For example, in the classical ABA paradigm (see e.g. [15]), a mixed sequence of sounds can be perceived as originating from one or two sources. If the different points in time in the sequence are considered as separate stimuli and subjected to gRSA, their distance varies in a way mimicking the perceived distance between/number of streams for a number of classical parameters, e.g. onset timing or frequency difference (preliminary simulation results, not shown).

A second, particularly interesting application of gRSA would be spatial hearing, where the set of spatial directions tested in the lab typically forms a sphere, i.e. where analogously to the Shepard tones, correct distance measurement hinges on knowing the shortest path on the sphere.

Lastly, as in vision, similarity analysis would be particulary useful for analyzing the grouping of complex stimuli, such as speech or music, where the physical geometry is in itself not easily defined, and deducing it from the neural responses is of even greater interest.

References

[1] S. Edelman: Representation is representation of similarities. Behavioral and Brain Sciences **21** (1998) 449–467.

[2] W. S. Torgerson: Multidimensional scaling: I. theory and method. Psychometrika **17** (1952) 401–419.

[3] R. N. Shepard: Attention and the metric structure of the stimulus space. Journal of mathematical psychology **1** (1964) 54–87.

[4] R. N. Shepard: Multidimensional scaling, tree-fitting, and clustering. Science (New York, N.Y.) **210** (Okt. 1980) 390–398.

[5] N. Kriegeskorte, R. A. Kievit: Representational geometry: integrating cognition, computation, and the brain. Trends in cognitive sciences **17** (2013) 401–412.

[6] N. Kriegeskorte, M. Mur, D. A. Ruff, R. Kiani, J. Bodurka, H. Esteky, K. Tanaka, P. A. Bandettini: Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron **60** (2008) 1126–1141.

[7] R. N. Shepard: Circularity in judgments of relative pitch. The Journal of the Acoustical Society of America **36** (1964) 2346–2353.

[8] J. K. Bizley, F. R. Nodal, I. Nelken, A. J. King: Functional organization of ferret auditory cortex. Cerebral cortex (New York, N.Y. : 1991) **15** (Okt. 2005) 1637–1653.

[9] B. Englitz, S. Akram, S. David, C. Chambers, D. Pressnitzer, D. Depireux, J. Fritz, S. A. Shamma: Putting the tritone paradox into context: insights from neural population decoding and human psychophysics. – In: Basic Aspects of Hearing. Springer, 2013, 157–164.

[10] R. N. Shepard: The analysis of proximities: multidimensional scaling with an unknown distance function. i. Psychometrika **27** (1962) 125–140.

[11] B. H. Repp: Spectral envelope and context effects in the tritone paradox. Perception **26** (1997) 645–665.

[12] M. C. van Wezel, W. A. Kosters, P. Van Der Putten, J. N. Kok: Nonmetric multidimensional scaling with neural networks. International Symposium on Intelligent Data Analysis, 2001, Springer, 145–155.

[13] D. E. Winkowski, P. O. Kanold: Laminar transformation of frequency organization in auditory cortex. Journal of Neuroscience **33** (2013) 1498–1508.

[14] A. S. Bregman, et al.: Auditory scene analysis. 1990.

[15] H. Gockel, B. C. Moore, R. P. Carlyon: Influence of rate of change of frequency on the overall pitch of frequency-modulated tones. The Journal of the Acoustical Society of America **109** (2001) 701–712.