NEUROCOMPUTING

ELSEVIER

Neurocomputing 32–33 (2000) 913–919

www.elsevier.com/locate/neucom

Modeling auditory cortical processing as an adaptive chirplet transform☆

Eduardo Mercado III*, Catherine E. Myers, Mark A. Gluck

Center for Molecular and Behavioral Neuroscience, Rutgers University, 197 University Ave., Newark, NJ 07102, USA

Accepted 13 January 2000

Abstract

Recent evidence suggests that (a) auditory cortical neurons are tuned to complex time-varying acoustic features, (b) auditory cortex consists of several fields that decompose sounds in parallel, (c) the metric for such decomposition varies across species, and (d) auditory cortical representations can be rapidly modulated. Past computational models of auditory cortical processing cannot capture such representational complexity. This paper proposes a novel framework in which auditory signal processing is characterized as an adaptive transformation from a one-dimensional space into an *n*-dimensional auditory parameter space. This transformation can be modeled as a chirplet transform implemented via a self-organizing neural network. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Neural; Wavelet; Unsupervised learning; Plasticity; Receptive field

1. Introduction

How networks of cortical neurons represent sound is poorly understood. Although electrophysiological studies have demonstrated that firing patterns in specific neural regions can often be predictably correlated with particular sound features, the underlying neural codes that give rise to such correlations remain unclear. Recently, there have been an increasing number of attempts to develop signal processing models of audition [16,26,27]. The motivation behind these efforts is the hope that current

☆This work was supported by a postdoctoral fellowship from APA/NIMH, and by the Rutgers MBRS program.

* Corresponding author. Tel.: + 1-973-353-1080-3226; fax: + 1-973-353-1272.

E-mail address: mercado@pavlov.rutgers.edu (E. Mercado III).

PII: S0925-2312(00)00260-5

computational techniques can provide insight into how neural circuits encode representations of acoustic events. In the current paper, we propose a heuristic model of auditory cortical processing based on recently described signal transformation techniques and self-organizing neural networks. This model encapsulates much of what is currently known about the response properties of auditory cortex. These properties include:

- Complex patterns of sound feature selectivity [8,16,22,36].
- Species-specific signal decomposition [28,31,33].
- Dynamic modulation of response characteristics [3,5,10–12,37].
- Systematic topography of cortical sensitivities [1,30].

Our intention is that the model be flexible enough that it can be used to describe how sounds are cortically encoded in any mammalian species.

Past computational models of auditory cortical processing have focused primarily on emulating neural sensitivities measured from individual neurons in a particular species. For example, Suga [32,33] described auditory processing in bats as parallel, hierarchical cross correlation. He modeled individual neurons as delay lines, multipliers, logical gates, and filters that decomposed incoming signals into functionally relevant temporal and/or spectral features. Processing in bats has also been modeled as (1) spectrographic cross correlation, followed by transformation into auditory parameters [29], (2) transformation from neural spike trains into a spatial array of delay sensitive units [25,26], and (3) binaural recognition of waveform envelopes [17]. Perhaps the most sophisticated models of auditory cortical processing developed to date characterize response sensitivities in ferrets [16,34,35]. Wang and Shamma [35] modeled auditory cortical neurons as perceptrons with weight vectors corresponding to the neural sensitivities of ferrets. These sensitivities were found to be analogous to a topographically organized wavelet transform. Attempts have also been made to model auditory processing in humans [4,7,18]. These models tend to focus on known psychophysical sensitivities (e.g., timbre and pitch perception) rather than electrophysiological response properties.

The standard approach to modeling auditory cortical processing has been to start with a general signal processing model (e.g., Fourier or wavelet transforms), and then to add on specialized processing components (e.g., matched filters) to reflect species-specific sensitivities. This approach is problematic because (1) the customization needed to describe cortical sensitivities in any given species can only be retrospectively determined (i.e., the models are not predictive), (2) most evidence suggests that auditory cortical sensitivities reflect the particular needs of individuals faced with species-specific ecological and biological constraints, rather than generic acoustic signal processing strategies, and (3) experience-dependent adaptations in auditory processing are not considered. A more flexible framework is needed to adequately characterize the full range of auditory cortical sensitivities observed in mammals. Our approach involves first finding a transform general enough to describe cortical signal decomposition across all mammals. This transform is then mapped onto an unsupervised neural network than can learn to efficiently code acoustic events that are of functional relevance to a particular species/individual.

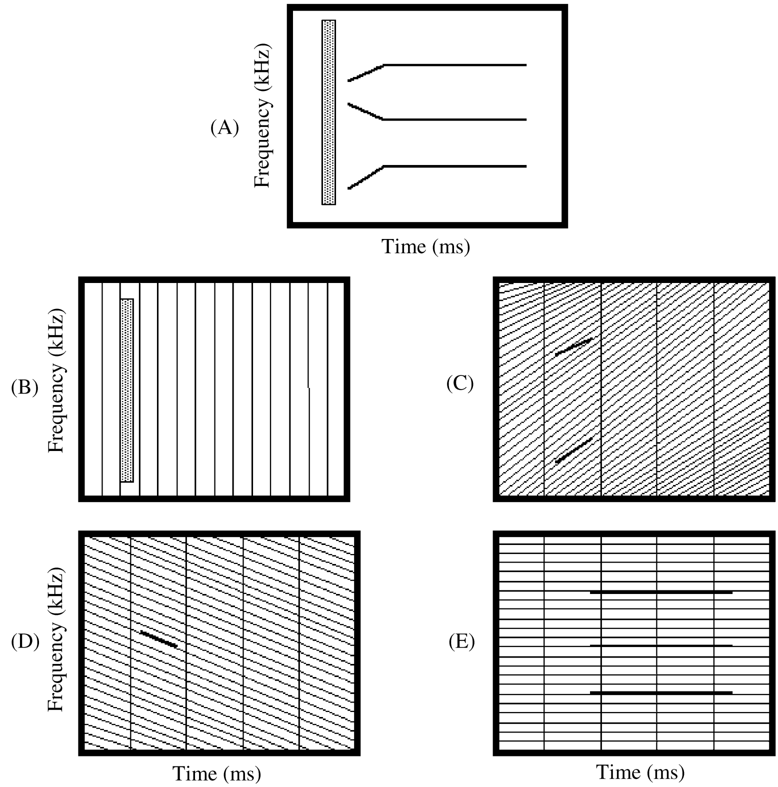

Fig. 1. An example of chirplet decomposition. (A) Idealized spectrogram of a spoken syllable (based on [32]). Dark regions of the spectrogram reflect higher spectral energy levels during short time intervals. The chirplet transform describes spectrograms such as (A) in terms of divisions of the time–frequency plane. For example, in (B) the initial broad band noise burst in the syllable (corresponding to a consonant) can be described as the third "slice" when the plane is vertically (i.e., temporally) segmented. In contrast, frequency-modulated components of the syllable (corresponding to the onset of a vowel), shown in (C) and (D) are better described in terms of divisions that segment the plane both vertically and diagonally. Finally, (E) shows that continuous frequency components are best characterized in terms of orthogonally segmented divisions. Chirplet spaces are defined based on the range of allowable divisions of the time–frequency plane. In the example above, dimensions of chirplet space correspond to the positions, sizes, and tilts of parallelograms covering the plane. Any syllable can be described in terms of a set of parallelograms that contain high concentrations of energy; each possible parallelogram corresponds to a point within the chirplet space. It is important to note that the shapes of segments are not limited to parallelograms. The basis functions chosen for the chirplet transform specify the geometry of segmentation.

2. Adaptive chirplets

A recently developed signal processing model, called the chirplet transform, appears to be well suited for our purposes. The chirplet transform subsumes both Fourier analysis and wavelet analysis (as well as several other classes of time-frequency analysis) as lower dimensional subspaces in the chirplet analysis space [19–21], providing a broad framework for mapping one-dimensional sound waveforms into an n-dimensional auditory parameter space. Fig. 1 illustrates the basic structure of the chirplet transform.

The chirplet transform retains the advantages offered by time frequency and wavelet transforms, and additionally provides a natural way for characterizing the different types of processing that have been described for different auditory fields (i.e., cortical regions with systematically related response sensitivities). Each auditory field can be viewed as a processor for decomposing sounds within a particular subspace of the complete auditory parameter space. In our framework, these fields correspond either to chirplet subspaces or to chirplet spaces generated by sets of functionally relevant basis functions. Chirplet spaces are highly overcomplete (redundant) because there are an infinite number of ways to segment a time-frequency plane. Because of this overcompleteness, the same acoustic feature may be encoded multiple times. Such multiplicative, overcomplete encoding corresponds well with the overlapping, parallel signal processing pathways observed in mammalian auditory cortex [1].

The flexibility of the chirplet transform comes at the price of loose constraints. We would like to focus only on the dimensions/spaces that are closest to those used in auditory cortical processing. However, for most species the relevant auditory parameters are unknown, making it difficult to choose either appropriate basis functions or dimensions. One way to address this issue is by developing adaptive models with constrained inputs and constrained learning abilities. The feature decompositions learned by these models can then be compared with those observed in cortex. For example, Olshausen and Field [23,24] have developed unsupervised learning algorithms that find linear codes for natural visual scenes, given the constraints that these codes are sparse and statistically independent. The codes generated by their algorithms decompose images in ways similar to simple cells in visual cortex and wavelet transforms. Applying these algorithms to natural acoustic scenes and/or species-typical vocalizations may provide insights into which chirplet spaces are most applicable.

Olshausen and Fields' [23,24] adaptive image decomposition techniques are limited in that they do not account for coding and recognition of patterns that have been translated, rotated, or scaled; such coding is intrinsic to the chirplet transform. Their approach also does not incorporate the topographic feature decomposition typical of cortical processing. Kohonen [13–15] has developed a neural network, called an adaptive subspace self-organizing map, that addresses these limitations. Individual units in the self-organizing map (which are themselves composed of multiple computational neurons) learn to represent sets of input patterns that fall within a particular subspace. This network can learn to encode simple transformations (e.g., translation), and organizes such transformation "detectors" topographically. Interestingly, signal decompositions learned by this neural network have also been found to be comparable to both wavelet and visual cortical decomposition [13].

In our framework, each unit in an adaptive subspace self-organizing map can be viewed as representing a dimension within a chirplet subspace (given a particular set of chirplet basis functions). Sets of units that share common bases can be considered to be analogous to a cortical field. Other wavelet-based self-organizing maps (e.g., see [6,9]) can also potentially be used to characterize auditory cortical parameter spaces.

The major obstacle blocking further development of this theoretical framework is the limited information available about the response properties of populations of auditory cortical neurons. Most attempts at characterizing the sensitivities of auditory cortex have looked at how individual neurons in primary auditory cortex respond to impoverished acoustic stimuli. Detailed descriptions of spectrotemporal response sensitivities have only recently begun to be reported [16,34]. Until such data are collected from a variety of species, it will be difficult to assess how effective our approach is at modeling encoding of acoustic events in mammalian cortices.

3. Conclusion

In this paper, we described a model of auditory cortical processing that more accurately reflects the complexity, variability, and flexibility seen in mammals. This model maps an overcomplete, acoustic signal decomposition onto a topographically organized, unsupervised neural network.

Our approach differs from previous approaches in that we start with an over-specified auditory parameter space, and attempt to reduce this space to reflect species-specific response sensitivities. Additionally, because our model is adaptive, it can be used to investigate changes in response sensitivities induced by experience (see also [2]).

The adaptive chirplet framework suggests new experimental directions for describing auditory cortical sensitivities. For example, measurements of experience-dependent changes in response sensitivities could be used to identify "paths of least resistance" in auditory parameter space. Such preferred trajectories could provide important clues about constraints on cortical sound decomposition and the chirplet bases/dimensions that best describe this process.

References

[1] L. Aitkin, The Auditory Cortex, Chapman and Hall, London, 1990.
[2] J.L. Armony, D. Servan-Schreiber, J.D. Cohen, J.E. LeDoux, An anatomically constrained neural network model of fear conditioning, Behav. Neurosci. 109 (1995) 246–257.
[3] J.S. Bakin, N.M. Weinberger, Induction of physiological memory in the cerebral cortex by stimulation of the nucleus basalis, Proc. Natl. Acad. Sci. USA 93 (1996) 11 219–11 224.
[4] J.J. Barucha, W.E. Mencl, Two issues in auditory cognition: self-organization of octave categories and pitch-invariant pattern recognition, Psychol. Sci. 7 (1996) 142–149.
[5] D.V. Buonomano, M.M. Merzenich, Cortical plasticity: from synapses to maps, Ann. Rev. Neurosci. 21 (1998) 149–186.
[6] M.M. Campos, G.A. Carpenter, WSOM: building adaptive wavelets with self-organizing maps, Proceedings of the 1998 IEEE WCCI, Anchorage, AK, 1998, pp. 763–767.
[7] P. Cosi, G. De Poli, G. Lauzzana, Auditory modelling and self-organizing neural networks for timbre classification, J. New Mus. Res. 23 (1996) 71–98.

[8] R.C. deCharms, D.T. Blake, M.M. Merzenich, Optimizing sound features for cortical neurons, Science 280 (1998) 1439–1443.
[9] Q.Q. Huynh, L.N. Cooper, N. Intrator, Classification of underwater mammals using feature extraction based on time-frequency analysis and BCM theory, IEEE Trans. Signal Process 46 (1998) 1202–1207.
[10] J.H. Kaas, Plasticity of sensory representations in the auditory and other systems of adult mammals, in: R.J. Salvi, D. Henderson (Eds.), Auditory System Plasticity and Regeneration, Thieme Medical Publishers, New York, 1996, pp. 213–223.
[11] M.P. Kilgard, M.M. Merzenich, Cortical map reorganization enabled by nucleus basalis activity, Science 279 (1998) 1714–1718.
[12] M.P. Kilgard, M.M. Merzenich, Plasticity of temporal information processing in the primary auditory cortex, Nat. Neurosci. 1 (1998) 727–731.
[13] T. Kohonen, Emergence of invariant-feature detectors in the adaptive-subspace self-organizing map, Biol. Cybernet. 75 (1996) 281–291.
[14] T. Kohonen, Self-Organizing Maps, Springer, Berlin, 1997.
[15] T. Kohonen, S. Kaski, H. Lappalainen, Self-organized formation of various invariant-feature filters in the adaptive subspace SOM, Neural Comput. 9 (1997) 1321–1344.
[16] N. Kowalski, D.A. Depireux, S.A. Shamma, Analysis of dynamic spectra in ferret auditory cortex: II, Prediction of unit responses to arbitrary dynamic spectra, J. Neurophys. 76 (1996) 3524–3534.
[17] R. Kuc, Biomimetic sonar recognizes objects using binaural information, J. Acoust. Soc. Am. 102 (1997) 689–696.
[18] M. Leman, Emergent properties of tonality functions by self-organization, J. New Mus. Res. 19 (1990) 85–106.
[19] S. Mann, S. Haykin, Chirplets and warblets: novel time-frequency methods, Electron. Lett. 28 (1992) 114–116.
[20] S. Mann, S. Haykin, The chirplet transform: Physical considerations, IEEE Trans. Signal Process. 43 (1995) 2745–2761.
[21] D. Mihovilovic, R.N. Bracewell, Adaptive chirplet representation of signals on time-frequency plane, Electron. Lett. 27 (1991) 1159–1161.
[22] I. Nelken, Y. Rotman, O.B. Yosef, Responses of auditory-cortex neurons to structural features of natural sounds, Nature 397 (1999) 154–157.
[23] B.A. Olshausen, D.J. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images, Nature 381 (1996) 607–609.
[24] B.A. Olshausen, D.J. Field, Sparse coding with an overcomplete basis set: a strategy employed by V1?, Vision Res. 37 (1997) 3311–3325.
[25] M.J. Palakal, U. Murthy, S.K. Chittajallu, D. Wong, Tonotopic representation of auditory responses using self-organizing maps, Math. Comput. Modelling 22 (1995) 7–21.
[26] M.J. Palakal, D. Wong, Cortical representation of spatiotemporal pattern of firing evoked by echolocation signals: population encoding of target features in real time, J. Acoust. Soc. Am. 106 (1999) 479–490.
[27] J.W. Pitton, K. Wang, B.-H. Juang, Time-frequency analysis and auditory modeling for automatic recognition of speech, Proc. IEEE 84 (1995) 1199–1214.
[28] G.D. Pollak, J.A. Winer, W.E. O'Neill, Perspectives on the functional organization of the mammalian auditory system: why bats are good models, in: A. Popper, R. Fay (Eds.), Hearing by Bats, Springer, New York, 1995, pp. 481–498.
[29] P.A. Saillant, J.A. Simmons, S.P. Dear, T.A. McMullen, A computational model of echo processing and acoustic imaging in frequency-modulated echolocating bats: the spectrogram correlation and transformation receiver, J. Acoust. Soc. Am. 94 (1993) 2691–2712.
[30] H. Scheich, Representational geometries of telencephalic auditory maps in birds and mammals, in: B. Finlay, G. Innocenti, H. Scheich (Eds.), The Neocortex: Ontogeny and Phylogeny, Plenum Press, New York, 1990, pp. 119–136.
[31] H. Scheich, Auditory cortex: comparative aspects of maps and plasticity, Curr. Opin. Neurobiol. 1 (1991) 236–247.

[32] N. Suga, Cortical computational maps for auditory imaging, Neural Networks 3 (1990) 3–21.

[33] N. Suga, Processing of auditory information carried by species-specific complex sounds, in: M.S. Gazzaniga (Ed.), The Cognitive Neurosciences, MIT Press, MA, 1995, pp. 295–313.

[34] H. Versnel, S.A. Shamma, Spectral-ripple representation of steady-state vowels in primary auditory cortex, J. Acoust. Soc. Am. 103 (1998) 2502–2514.

[35] K. Wang, S.A. Shamma, Auditory analysis of spectro-temporal information in acoustic signals, IEEE Eng. Med. Biol. 14 (1994) 186–194.

[36] X. Wang, M.M. Merzenich, R. Beitel, C.E. Schreiner, Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics, J. Neurophys. 74 (1995) 2685–2706.

[37] N.M. Weinberger, Dynamic regulation of receptive fields and maps in the adult sensory cortex, Ann. Rev. Neurosci. 18 (1995) 129–158.

Eduardo Mercado III earned his B.E. in Computer Engineering from Georgia Tech in 1993. From his studies at the Kewalo Basin Marine Mammal Laboratory (University of Hawaii) he took an M.A. in Psychology (An acoustical analysis of humpback whale song units, 1995) an M.Sc. in Electrical Engineering (Computational models of sound production and reception in the humpback whale, 1998), and a Ph.D. in Psychology (Humpback whale bioacoustics, 1998). Since 1998, he has been working as a Postdoctoral Fellow in the Center for Molecular and Behavioral Neuroscience at Rutgers University. His current work focuses on comparative and computational models of corticohippocampal processing of acoustic events.

Catherine E. Myers is a Research Assistant Professor in the Department of Psychology at Rutgers University. She received her B.S. in Cognitive Science from the University of Delaware in 1987, and her Ph.D. in Neural Neworks from the University of London in 1990. Her current interests and work focus on modeling the contribution of the hippocampal region to learning and memory in humans and animals, using connectionist network models. The data she focuses on are drawn from associative learning (especially classical and operant conditioning) in animals and humans with intact or damaged hippocampal regions.

Mark A. Gluck is an Associate Professor in the Center of Molecular and Behavioral Neuroscience at Rutgers University. He graduated from Harvard University in 1982 (B.A. Psychology and Computer Science) and obtained his Ph.D. at Stanford University in 1987 (Cognitive Psychology). Currently, he combines his training in human and animal psychology with neural-network analyses in order to identify, model, and empirically evaluate fundamental components of both animal and human memory systems, with a special emphasis on the functional role of the hippocampus in learning and memory.