Non-isomorphism in efficient coding of complex sound properties

Christian E. Stilp and Keith R. Kluender

*Department of Psychology, University of Wisconsin–Madison, 1202 West Johnson Street,
Madison, Wisconsin 53706
christian.stilp@gmail.com, krkluend@wisc.edu*

Abstract: To the extent that sensorineural systems are efficient, stimulus redundancy should be captured in ways that optimize information transmission. Consistent with this principle, neural representations of sounds have been proposed to become “non-isomorphic,” increasingly abstract and decreasingly resembling the original (redundant) input. Here, non-isomorphism is tested in perceptual learning using AXB discrimination of novel sounds with two highly correlated complex acoustic properties and a randomly varying third dimension. Discrimination of sounds obeying the correlation became superior to that of sounds violating it despite widely varying physical acoustic properties, suggesting non-isomorphic representation of stimulus redundancy.

© 2011 Acoustical Society of America

PACS numbers: 43.66.Ba, 43.66.Lj [QJF]

Date Received: July 28, 2011 Date Accepted: September 6, 2011

1. Introduction

Much of the stimulation available to perceivers is redundant because some sensory attributes can be predicted from other attributes concurrently, successively, or as a consequence of experience with a structured environment. To the extent that sensorineural systems are efficient, redundancy should be extracted to optimize transmission of information. For example, [Chechik and colleagues \(2006\)](#) provided physiological evidence that neural responses at successive stages of processing in the auditory system become increasingly independent from one another. Capitalizing on regularities across stimuli has a host of perceptual benefits: uncertainty is reduced, neural coding becomes more efficient, sensitivity to stimulus associations is heightened, and interactions with the environment become informed through learning. While these principles of “efficient coding” ([Attneave, 1954](#); [Barlow, 1961](#); [Simoncelli, 2003](#)) have proven productive for sensory and computational neuroscience, perceptual evidence has been limited.

[Stilp and colleagues \(2010\)](#) provided the first direct behavioral evidence for efficient auditory perceptual learning. Listeners heard novel highly controlled sounds that varied along two physically independent complex acoustic dimensions: attack/decay (AD) and spectral shape (SS). All steps between stimuli along both dimensions were psychoacoustically equivalent. Listeners were presented a set of sounds across which AD and SS were highly correlated. Early in testing, robust stimulus covariance was encoded efficiently such that discrimination of sounds obeying the correlation is maintained but is significantly impaired for single dimensions AD and SS and for sounds violating the correlation (i.e., varying in both AD and SS but in an orthogonal manner). These differences in discrimination are not observed when dimensions are weakly correlated ([Stilp et al., 2010](#)) or when greater evidence is provided for an orthogonal dimension ([Stilp and Kluender, 2010](#)). This perceptual reorganization cannot be explained by independent weighting of acoustic dimensions (AD, SS), as changes in discriminability can only be attributed to the correlation or covariance orthogonal to it.

For any process that efficiently captures redundancy, it is necessarily true that neural representations must become decreasingly like the stimulus. This is because

systematically covarying stimulus properties collapse into more efficient representations at the expense of separate redundant properties. Consistent with this principle, Wang (2007) describes “non-isomorphic” transformations that occur progressively along the ascending auditory pathway, making neural representations “further away from physical (acoustical) structures of sounds, but presumably closer to internal representations underlying perception” (p. 92). Examples of non-isomorphic representations in auditory cortex include encoding spectral shape across varying absolute frequencies (Barbour and Wang, 2003), gross representation of rapid change in click trains with short inter-click intervals versus phase-locking to trains with slower inter-click intervals (Lu and Wang, 2000; Lu *et al.*, 2001), pitch versus individual frequency components (Bendor and Wang, 2005, 2006), and different components of auditory scenes (Nelken and Bar-Yosef, 2008).

Such non-isomorphic transformations may be similar to the loss of acoustic dimensions (AD, SS) as more efficient dimensions better capture perceptual performance (Stilp *et al.*, 2010). However, all acoustic variability in the experiments of Stilp *et al.* served to directly define or violate correlation between changes along two dimensions. Natural sounds are more acoustically complex, varying along many acoustic dimensions. In many or most cases, changes along multiple dimensions are not all correlated. The extent to which efficient coding persists in more naturalistic circumstances, when some acoustic dimensions are correlated while others vary in random or irrelevant ways, is unclear. The present experiment formally tests non-isomorphic representation of stimulus redundancy in auditory perceptual learning. To the extent that efficient coding of correlation between two attributes is non-isomorphic, irrelevant variation in a third attribute should not alter patterns of performance, and listeners should exhibit superior discrimination of sound pairs obeying the correlation versus those violating it.

2. Methods

2.1 Participants

Forty undergraduate students from the University of Wisconsin participated in the experiment. All reported no known hearing impairments. They were compensated for their time with extra credit in an introductory psychology course.

2.2 Stimuli

All stimuli are novel complex sounds described in detail by Stilp *et al.* (2010). Briefly, three pitch pulses from samples of a French horn and a tenor saxophone playing the same note (Opolko and Wapnick, 1989) were iterated to 500-ms duration and RMS-matched in amplitude. Samples were then edited to vary along one of two complex acoustic dimensions: attack/decay (AD) or spectral shape (SS), dimensions that are in principle relatively independent both perceptually and in early neural encoding (Caclin *et al.*, 2006). AD was defined as the amplitude envelope of the stimulus which was set to zero at stimulus onset and offset, with linear ramps from onset to peak and back to offset without any steady state. SS was manipulated by adding instrument endpoints in different proportions, ranging from 0.2 to 0.8 for each instrument and always summing to 1.0. Differences between mixtures were derived by calculating Euclidean distances between ERB-scaled spectra (Glasberg and Moore, 1990) that had been processed by a bank of auditory filters (Patterson *et al.*, 1982). AD and SS series were then exhaustively adjusted across hundreds of participants until every pair of sounds separated by three stimulus steps (out of 18 steps total) was equally discriminable to every other pair within and across stimulus series ($\approx 65\%$ correct for changes along one dimension, $\approx 70\%$ for changes along both dimensions; see Stilp *et al.*, 2010 for details).

A third acoustic dimension, vibrato, was developed through a separate series of norming experiments. Vibrato is introduced through sinusoidal modulation of frequency components. Thus, mean fundamental frequency stays constant while

fixed-depth frequency excursions modulate. Critically, manipulating vibrato does not alter global AD or SS properties. Strictly speaking, vibrato causes the full spectrum to shift up and down in absolute frequency while maintaining constant spectral shape. Thus, vibrato-induced changes are similar to encoding spectral shape across varying absolute frequencies in physiological studies of cortical encoding (Barbour and Wang, 2003). Vibrato was varied in 18 nearly logarithmic steps from 7.5–19 Hz, with step sizes normed in pilot studies to share equivalent JND spacing as achieved for AD and SS series.

All three acoustic dimensions (AD, SS, vibrato) were fully crossed to generate a stimulus cube of 5 832 sounds, a small subset of which was used in the experiment. Similar to the design tested by Stilp and Kluender (2010), AD and SS were near-perfectly correlated with each other [$r = \pm 0.98$, calculated using nominal values from 1 to 18 to represent AD and SS values; Fig. 1(a)]. One listener group ($n = 20$) was presented the positive correlation between AD and SS while the other half heard the negative correlation between dimensions. Consequently, one group's Consistent dimension served as the other group's Orthogonal dimension and vice versa. AD and SS were varied to generate 16 stimulus pairs, each separated by three stimulus steps: 15 pairs supporting robust correlation (Consistent condition), and one pair directly violating it (Orthogonal condition).

Sixteen values of vibrato (varying from 7.5–17.2 Hz) were randomly assigned to each of the 16 AD/SS stimulus pairs [Fig. 1(b)], separately for each listener group.

Fig. 1. (Color online) Stimuli and results from the experiment. (a) Robust correlation between AD and SS as tested by Stilp and Kluender (2010). Eighteen sounds lie on the main diagonal and support the correlation (Consistent condition; triangles) while two sounds lay on the opposing diagonal, directly violating the correlation (Orthogonal condition; squares). (b) Three-dimensional stimulus cube, with circles depicting all sounds presented in the experiment. Variability in vibrato cues is evident, but AD and SS maintain their robust correlation with one another (triangles and squares, collapsed across vibrato values). (c) AXB discrimination accuracy as a function of testing block number. Error bars depict standard error of the mean. Asterisk indicates significant difference ($p < 0.05$) as assessed by a paired-sample t -test.

Sound pairs were then arranged into AXB triads (64 per group) with 250-ms ISIs. Thus, while vibrato varied from trial to trial (from one pair to the next), each sound in an AXB triad featured the same vibrato so discrimination was on the basis of variation in AD and SS only.

2.3 Procedure

Sounds were upsampled to 48 828 Hz, D/A converted (Tucker-Davis Technology RP2.1), amplified (TDT HB4), and presented diotically over circumaural headphones (Beyer Dynamic DT-150) at 72 dBA. Between one and three individuals participated concurrently in single-subject soundproof booths. Each participant heard trials in a different randomized order. Trials were presented twice in each of three blocks, 128 trials per block for a total of 384 responses per listener. No feedback was provided. Listeners were given the opportunity to take a short break between testing blocks. The experiment lasted approximately 30 min.

3. Results

Performance data are presented in Fig. 1(c), with discrimination accuracy (proportion correct) on the ordinate and testing block on the abscissa. Given that learning experiments of this type are expected to reveal changes in discriminability across testing blocks, omnibus analysis of variance tests are likely to result in Type II error. Consequently, to retain sensitivity to differences in discriminability across conditions at different phases of the experiment, results are analyzed using planned-comparison paired-sample two-tailed *t*-tests. Discrimination of Consistent sound pairs was not significantly different from that of the Orthogonal pair in the first (Consistent mean = 0.61, s.e. = 0.01; Orthogonal mean = 0.59, s.e. = 0.03; $t_{39} = 0.62$, *n.s.*) or second testing block (Consistent mean = 0.63, s.e. = 0.02; Orthogonal mean = 0.63, s.e. = 0.03; $t_{39} = 0.02$, *n.s.*). Consequent to further experience with the correlation between AD and SS, discrimination did significantly differ in the third block with Consistent sound pairs discriminated more accurately than the Orthogonal pair (Consistent mean = 0.65, s.e. = 0.01; Orthogonal mean = 0.59, s.e. = 0.03; $t_{39} = 2.16$, $p < 0.05$).

4. Discussion

Efficient coding of redundant acoustic dimensions persists in the face of uncorrelated acoustic variation of comparable magnitude. Despite random variability in vibrato rate from trial to trial, listeners still came to perform significantly better discriminating sound pairs obeying the correlation (Consistent) relative to those violating it (Orthogonal). Consistent with previous findings (Stilp *et al.*, 2010; Stilp and Kluender, 2010), discriminability was again predicted by patterns of covariance among acoustic properties rather than the acoustic properties themselves. Efficient coding develops even in the presence of substantial random variability along a third acoustic dimension, as predicted by abstract (i.e., non-isomorphic) representation of stimulus redundancy.

Present results provide a significant extension of earlier observations. In previous studies where all stimulus change was along only two dimensions (Stilp *et al.*, 2010; Stilp and Kluender, 2010), discriminating Consistent sound pairs significantly better than Orthogonal pairs occurred early in testing. By contrast, these differences in discriminability were delayed, but not diminished, in the face of trial-to-trial variation in a third uncorrelated dimension. Greater experience with the statistical structure of the stimuli was required to discount the uncorrelated dimension and to efficiently code stimuli.

For the carefully linearized stimulus sets (equal JND steps) created for these experiments, performance can be characterized by principal component analysis-type operations (Stilp *et al.*, 2010) that capture derived non-isomorphic dimensions. Other forms of stimulus recoding can also result in non-isomorphism, and it is likely that different or additional processes including nonlinear transformations would be required for more natural stimulus sets.

It should be noted that the ability to extract efficient representations in the face of uncorrelated variability does not preclude the coexistence of representations that more faithfully encode stimulus properties (Nelken and Bar-Yosef, 2008). For example, while Barbour and Wang (2003) report neural sensitivity in primary auditory cortex to levels of spectral contrast (non-isomorphic), several other reports document neural encoding of the gross frequency characteristics of a stimulus (isomorphic; e.g., Wang *et al.*, 1995). More confident speculation concerning underlying sensorineural processing responsible for the present findings must await additional behavioral and physiological experiments. Nevertheless, performance reported here cannot be explained by representation of physical acoustic dimensions, but only by representation of the covariance between them.

Results presented here may provide insights into models of perceptual organization for complex sounds such as speech. While the novel sounds employed here varied only along three acoustic dimensions (one of which varied randomly), patterns of covariance naturally scale to high-dimensional feature spaces. In complex natural stimuli such as speech, multiple forms of stimulus attribute redundancy exist concurrently and successively (e.g., Delattre *et al.*, 1955; Kluender *et al.*, 2011; Lisker, 1978; Repp, 1982; Sussman *et al.*, 1991, 1998). To the extent that patterns of covariance among acoustic attributes in natural sounds are efficiently coded, these non-isomorphic representations may inform how the auditory system exploits different patterns of redundancy to learn to distinguish different speech sounds. For example, extraction of relational properties across variations consequent to coarticulation (e.g., locus equations, Sussman *et al.*, 1991, 1998) or anatomy (scaling of formant frequencies across changes in vocal tract length across talkers, Kluender *et al.*, 2011) are the most direct speech analogs to non-isomorphism demonstrated here. In related studies employing fMRI, Okada *et al.* (2010) report that responses in bilateral posterior superior temporal sulcus were sensitive to phonemic variability (intelligibility) of speech sounds, but not to acoustic variability. These and other examples support the notion that high-level auditory processing captures abstract characteristics of complex stimuli. The present findings reveal that such an efficient, non-isomorphic representation can have profound effects on perceptual organization and stimulus discriminability even in the case of considerable irrelevant variability.

Acknowledgments

The authors wish to thank Nora Brand and Anna Joy Tan for assistance in conducting this experiment. This research was funded by grants from the National Institutes on Deafness and Other Communicative Disorders to C.E.S. (Grant No. F31 DC009532) and K.R.K. (Grant No. RC1 DC010601).

References and links

- Attneave, F. (1954). "Some informational aspects of visual perception," *Psychol. Rev.* **61**, 183–193.
- Barbour, D. L., and Wang, X. (2003). "Contrast tuning in auditory cortex," *Science* **299**, 1073–1075.
- Barlow, H. B. (1961). "Possible principles underlying the transformations of sensory messages," in *Sensory Communication*, edited by W. A. Rosenblith (MIT Press, Cambridge), pp. 53–85.
- Bendor, D., and Wang, X. (2005). "The neuronal representation of pitch in primary auditory cortex," *Nature (London)* **436**(7054), 1161–1165.
- Bendor, D., and Wang, X. (2006). "Cortical representations of pitch in monkeys and humans," *Curr. Opin. Neurobiol.* **16**, 391–399.
- Caclin, A., Brattico, E., Tervaniemi, M., Näätänen, R., Morlet, D., Giard, M.-H., and McAdams, S. (2006). "Separate neural processing of timbre dimensions in auditory sensory memory," *J. Cogn. Neurosci.* **18**, 1959–1972.
- Chechik, G., Anderson, M. J., Bar-Yosef, O., Young, E. D., Tishby, N., and Nelken, I. (2006). "Reduction of information redundancy in the ascending auditory pathway," *Neuron* **51**, 359–368.
- Delattre, P. C., Liberman, A. M., and Cooper, F. S. (1955). "Acoustic loci and transitional cues for consonants," *J. Acoust. Soc. Am.* **27**(4), 769–773.
- Glasberg, B. R., and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.* **47**, 103–138.

- Kluender, K. R., Stilp, C. E., and Kiefte, M. (2011). "Perception of vowel sounds within a biologically realistic model of efficient coding," in *Vowel Inherent Spectral Change*, edited by G. Morrison and P. Assmann (in press).
- Lisker, L. (1978). "Rapid versus ravid: A catalogue of acoustical features that may cue the distinction," Haskins Laboratories Status Report on Speech Research, SR-54, pp. 127–132.
- Lu, T., and Wang, X. (2000). "Temporal discharge patterns evoked by rapid sequences of wide- and narrow-band clicks in the primary auditory cortex of cat," *J. Neurophysiol.* **84**, 236–246.
- Lu, T., Liang, L., and Wang, X. (2001). "Temporal and rate representations of time-varying signals in the auditory cortex of awake primates," *Nat. Neurosci.* **4**, 1131–1138.
- Nelken, I., and Bar-Yosef, O. (2008). "Neurons and objects: The case of auditory cortex," *Front. Neurosci.* **2**(1), 107–113.
- Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.-H., Saberi, K., Serences, J. T., and Hickock, G. (2010). "Hierarchical organization of human auditory cortex: Evidence from acoustic invariance in the response to intelligible speech," *Cerebral Cortex* **20**(10), 2486–2495.
- Opolko, F., and Wapnick, J. (1989). *McGill University Master Samples User's Manual* (McGill University, Faculty of Music, Montreal).
- Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, D. (1982). "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," *J. Acoust. Soc. Am.* **72**, 1788–1803.
- Repp, B. H. (1982). "Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception," *Psychol. Bull.* **92**, 81–110.
- Simoncelli, E. P. (2003). "Vision and the statistics of the visual environment," *Curr. Opin. Neurobiol.* **13**, 144–149.
- Stilp, C. E., and Kluender, K. R. (2010). "Efficient coding of attenuated correlation among complex acoustic dimensions," *J. Acoust. Soc. Am.* **128**, 2455.
- Stilp, C. E., Rogers, T. T., and Kluender, K. R. (2010). "Rapid efficient coding of correlated complex acoustic properties," *Proc. Natl. Acad. Sci.* **107**(50), 21914–21919.
- Sussman, H. M., McCaffrey, H. A., and Matthews, S. A. (1991). "An investigation of locus equations as a source of relational invariance for stop place categorization," *J. Acoust. Soc. Am.* **90**, 1309–1325.
- Sussman, H. M., Fruchter, D., Hilbert, J., and Sirosh, J. (1998). "Linear correlates in the speech signal: The orderly output constraint," *Behav. Brain Sci.* **21**, 241–259.
- Wang, X. (2007). "Neural coding strategies in auditory cortex," *Hear. Res.* **229**, 81–93.
- Wang, X., Merzenich, M. M., Beitel, R., and Schreiner, C. E. (1995). "Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: Temporal and spectral characteristics," *J. Neurophysiol.* **74**, 2685–2706.