

Published in final edited form as:

Psychon Bull Rev. 2013 August ; 20(4): 740–746. doi:10.3758/s13423-013-0399-y.

Auditory rhythms are systemically associated with spatial frequency and density information in visual scenes

Aleksandra Sherman¹, Marcia Grabowecky^{1,2}, and Satoru Suzuki^{1,2}

¹Department of Psychology, Northwestern University, Evanston, IL

²Interdepartmental Neuroscience Program, Northwestern University, Evanston, IL

Abstract

A variety of perceptual correspondences between auditory and visual features have been reported, but few studies have investigated how rhythm, an auditory feature defined purely by dynamics relevant to speech and music, interacts with visual features. Here, we demonstrate a novel crossmodal association between auditory rhythm and visual clutter. Participants were shown a variety of visual scenes from diverse categories and were asked to report the auditory rhythm that perceptually matched each scene by adjusting the rate of amplitude modulation (AM) of a sound. Participants matched each scene to a specific AM rate with surprising consistency. A spatial-frequency analysis showed that scenes with larger contrast energy in midrange spatial frequencies were matched to faster AM rates. Bandpass-filtering the scenes indicated that large contrast energy in this spatial-frequency range is associated with an abundance of object boundaries and contours, suggesting that participants matched more cluttered scenes to faster AM rates. Consistent with this hypothesis, AM-rate matches were strongly correlated with perceived clutter. Additional results indicate that both AM-rate matches and perceived clutter depend on object-based (cycles per object) rather than retinal (cycles per degree of visual angle) spatial frequency. Taken together, these results suggest a systematic crossmodal association between auditory rhythm, representing density in the temporal domain, and visual clutter, representing object-based density in the spatial domain. This association may allow the use of auditory rhythm to influence how visual clutter is perceived and attended.

Keywords

Crossmodal; multisensory integration; spatial frequency; amplitude modulation rate; natural scenes; density; visual clutter

Introduction

Previous research has demonstrated a variety of perceptual correspondences between auditory and visual features. Most of these associations are based on auditory loudness mapping to visual brightness, auditory pitch (or pitch change) mapping to visual lightness, elevation, and size, and auditory timbre (often conveyed by speech sounds) mapping to sharpness/smoothness of visual contours or shapes (e.g., Bernstein & Edelman, 1971; Marks, 1987; Evans & Treisman, 2010; Mossbridge, Grabowecky, Suzuki, 2011; Ramachandran & Hubbard, 2001; Sweeny, Guzman-Martinez, Ortega, Grabowecky, & Suzuki, 2012).

Few studies have investigated how rhythm, an auditory feature defined purely by dynamics, may interact with visual features. Rhythm is a fundamental auditory feature coded in the auditory cortex, plays an integral role in providing information about objects and scenes (Liang, L., Lu, T., and Wang, X., 2002; Schreiner, C.E., and Urbas, J.V., 1986; 1998), and conveys affective and linguistic information in music and speech (e.g. Juslin & Laukka, 2003; Scherer, 1986; Bhatara, Tirovolas, Duan, Levy, & Levitin, 2011). Intuitively, a faster auditory rhythm is associated with visual properties that imply rapid dynamics. Consistent with this idea, Shintel and Nusbaum (2007) showed that listening to a verbal description of an object spoken at an atypically rapid rate speeded recognition of a subsequently presented picture when the picture depicted an object in motion relative to when it depicted the same object at rest. This suggests that auditory rhythm can interact with perception of visual dynamics.

Recently, Guzman-Martinez and colleagues (2012) have demonstrated that auditory rhythm is also associated with visual spatial frequency, a fundamental visual feature initially coded in the primary visual cortex (De Valois, Albrecht, & Thorell, 1982; Geisler & Albrecht, 1997) that is relevant for perceiving textures, objects, hierarchical structures, and scenes (Landy & Graham, 2004; Schyns & Oliva, 1994; Shulman, Sullivan, Gish, & Sakoda, 1986; Sowden & Schyns, 2006). They used a basic form of auditory rhythm conveyed by an amplitude-modulated (AM) sound (a white noise whose intensity is modulated at a fixed rate) and a basic form of visual spatial frequency conveyed by a Gabor patch (a repetitive grating-like pattern whose luminance is modulated at a fixed spatial frequency). They found that participants matched faster AM rates to higher spatial frequencies in an approximately linear relationship. This crossmodal relationship is absolute (rather than relative) in that it is equivalent whether each participant found an auditory match to only one Gabor patch, or found auditory matches to multiple Gabor patches of different spatial frequencies. The relationship is perceptual in that it is not based on general magnitude matching in an abstract numerical representation or matching the number of “bars” in Gabor patches to AM rates. It was further shown that the relationship is functionally relevant in that an AM sound can guide attention to a Gabor patch with the corresponding spatial frequency. These results suggest a fundamental relationship between the auditory processing of rhythm (AM rate) and the visual processing of spatial frequency.

Although it is necessary to characterize a crossmodal relationship using simplified visual stimuli, it is also important to understand how the relationship is relevant to perception in the real world. In the natural environment, we encounter complex scenes that are characterized by many spatial frequencies. It has been shown that responses of spatial-frequency tuned neurons to natural scenes are not readily predictable from their responses to Gabor patches (e.g., Olshausen & Field, 2006). In the current study, we thus investigated how the basic perceptual relationship between auditory rhythm and isolated visual spatial frequencies generalized to a perceptual relationship between auditory rhythm and natural scenes composed of multiple spatial-frequency components. This investigation will also elucidate how auditory rhythm may systematically influence processing of complex visual scenes.

Experiment 1

We first determined whether people would consistently match a variety of complex visual scenes to specific auditory AM rates. Namely, we asked, does a natural scene have an implied auditory rhythm? For example, a cluttered indoor scene may be matched to a faster AM rate than a less cluttered indoor scene, an urban scene to a faster AM rate than a beach scene, a mountain scene to a slower AM rate than a forest scene, and so on. Indeed, we found that people consistently matched each scene to a specific AM rate. We then analyzed

the spatial-frequency components of the images to investigate the source of this crossmodal association.

Methods

Participants—The participants in all experiments were Northwestern University undergraduate students, who gave informed consent to participate for partial course credit, had normal or corrected-to-normal visual acuity and normal hearing, and were tested individually in a dimly lit room. A group of 20 (9 female) students participated in Experiment 1.

Stimuli and Procedures—Participants determined auditory-AM-rate matches to 24 natural scenes (see Appendix) and three Gabor patches (0.50, 2.20, and 4.50 c/cm in physical spatial frequency, corresponding to 0.75, 3.30, and 6.79 c/d in retinal spatial frequency); see Figure 1 for trial information. All images were randomly presented 3 times, totaling 81 trials. Participants were given three practice trials prior to the experiment, in which they determined AM-rate matches to geometric patterns. Images were displayed full-screen on a 22-inch color CRT monitor (1152 × 870 pixels, 85Hz). An integrated head and chin rest was used to stabilize the viewing distance at 84 cm.

After auditory-visual matching trials were completed, participants determined whether each scene (not including the three Gabor patches) was dense or sparse with a forced-choice response. The image was slightly smaller (20.5° by 16.0° of visual angle) to present the choice words “dense” and “sparse” below the image. Each scene was randomly presented in two separate blocks. Left/right placement of the two words was counterbalanced (e.g. if the word “dense” appeared on the left in the first block, it appeared on the right in the second block, or vice versa). This provided a measure of perceived clutter.

The experiment was controlled using MATLAB software with Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997).

Results

Participants matched specific auditory AM rates to the 24 scenes from diverse categories (nature, urban, and indoor) with surprising consistency (Figure 2, black circles), $F(23, 437) = 17.43$, $p < .0001$ (main effect of images). The AM-rate matches to the intermixed Gabor patches were similar to those reported by Guzman-Martinez et al. (2012) in that there was no main effect of experiment ($F[1, 25] = 2.80$, *n.s.*), though there was a marginal interaction between experiment and spatial frequency ($F[2, 50] = 2.81$, $p = .07$) because our participants assigned slightly lower AM rates to the highest spatial-frequency Gabor patch (see Table 1 for the AM-rate matches in the two studies). We replicated a robust linear relationship between spatial frequency and AM rate ($t[19] = 6.10$, $p < .0001$, for linear contrast). The fact that our participants, who saw complex scenes as well as Gabor patches, and Guzman-Martinez et al.'s participants, who only saw Gabor patches, similarly matched Gabor patches of different spatial frequencies to specific AM rates, suggests that people tend to use a consistent strategy to match AM rates to both simple Gabor patches and complex scenes. Because Gabor patches primarily carry single spatial frequencies, we hypothesized that our participants used spatial frequency information to match AM rates to visual scenes.

To evaluate this hypothesis, we applied a 2D Fourier transform to each scene, and obtained its spatial-frequency profile with respect to 12 spatial-frequency bins ranging from 0.05 to 12.8 cycles per degree (c/d) (see Table 2).

For each participant, we computed the correlation between AM-rate matches and contrast energy for each spatial frequency bin. For example, if AM-rate matches were slower for scenes with larger energy in lower-spatial-frequency components, the correlations would be *negative* for lower-spatial-frequency bins. If AM-rate matches were faster for scenes with larger energy in higher-spatial-frequency components, the correlations would be *positive* for higher-spatial-frequency bins. Outlier images were removed from each correlation (across observers) using a 95% confidence ellipse (5% of images were removed on average). The average correlation coefficient, r , is shown as a function of spatial frequency bin in Figure 3 (black curve). The function is peaked. That is, we did not obtain a simple crossmodal relationship where the contrast energy in higher spatial frequency components drives faster AM-rate matches. Instead, the result suggests that the faster AM-rate matches are driven by the energy in the specific midrange spatial frequencies (0.3–1.25 c/d).

In order to gain insight into why the mid-range spatial frequencies are strongly associated with faster AM-rate matches, we filtered each image within this window of spatial frequency (0.3–1.25 c/d). Representative examples are shown in Figure 4. An inspection of these images suggests that scenes with stronger contrast energy in this spatial frequency bin tend to have more object boundaries and contours (e.g., the top image in Figure 4), whereas scenes with weaker contrast energy in this spatial frequency bin tend to have fewer object boundaries (e.g., the bottom image in Figure 4). This may suggest that AM-rate matches to visual scenes are based on the numerosity of object boundaries and contours. Consistent with this idea, we found a strong aggregate correlation between the average perceived clutter rating and average AM-rate match across the 24 scenes ($r = 0.62$, $t[22] = 3.68$, $p = 0.001$). A significantly positive correlation was also when it was computed separately for each participant, $t(19) = 3.89$, $p = .001$. This supports the hypothesis that larger contrast energy in the mid-range spatial frequencies drives a faster AM-rate match because it makes a scene appear more cluttered.

To determine whether spatial frequency information contributed to AM-rate matches over and above perceived clutter, we computed the correlation between the clutter rating and the contrast energy in each spatial frequency bin. If AM-rate matches are completely mediated by perceived clutter, the mid-range spatial frequencies that strongly drive faster AM-rate matches should also strongly drive higher clutter ratings. As can be seen in Figure 3 (gray curve), although the functions for perceived clutter and AM-rate matches are both broadly elevated within a similar range of spatial frequencies, the peak¹ occurs at a significantly higher spatial frequency for perceived clutter ($M=1.30$ c/d, $SD=1.93$) than for AM-rate matches ($M=0.44$ c/d, $SD=4.09$), $t(17) = 3.62$, $p < .01$. Thus, although spatial frequency and perceived clutter similarly contribute to AM-rate matches, perceived clutter depends on a slightly higher range of spatial frequencies than does the crossmodal association. This suggests that spatial frequency profiles of a visual scene contribute to AM-rate matches over and above perceived clutter.

Experiment 2

The goal of this experiment was to determine whether the spatial-frequency-mediated crossmodal association between visual scenes and auditory AM rate was based on retinal, physical, or object-based spatial frequency. In the case of single-spatial-frequency texture patches (Gabor patches), the association is based on physical spatial frequency (Guzman-Martinez et al., 2012). For texture perception, physical spatial frequency is informative

¹Peak locations were estimated by fitting each participant's tuning functions with quadratic polynomials. Two participants were excluded from the analysis (also in Experiment 2, coincidentally) because their peak locations deviated from the mean by greater than three standard deviations.

because it allows anticipation of felt texture prior to touching a surface. For scene perception, however, object-based spatial frequency (number of cycles per object) would be particularly informative because it conveys information about object structure and scene features irrespective of viewing distance or scaling of the photographs (e.g. Parish & Sperling, 1991; Sowden & Schyns, 2006). Thus, it is possible that AM-rate matches to natural scenes may be based on object-based (rather than physical) spatial frequency.

We tested this hypothesis by reducing the size of the images by half, thus doubling the physical/retinal spatial frequencies (physical and retinal spatial frequencies are indistinguishable at a fixed viewing distance) without affecting object-based spatial frequencies. If the crossmodal matches depend on physical/retinal spatial frequencies, AM-rate matches to the individual images should change, but the critical spatial frequencies (that are strongly correlated with AM-rate matches) should remain the same in cycles per degree. In contrast, if the crossmodal matches depend on object-based spatial frequencies, AM-rate matches to the individual images should remain the same, but the critical spatial frequency should double in cycles per degree because an identical object-based spatial frequency corresponds to a doubled physical/retinal spatial frequency when an image size is halved.

Methods

Participants—A new group of 14 undergraduate students (9 female) participated.

Stimuli and Procedures—The stimuli and procedures were the same as Experiment 1, except that image size was halved (to 13.2° by 9.87° of visual angle).

Results

Auditory AM-rate matches to the 24 scenes (Figure 2, gray squares) remained statistically equivalent to those in Experiment 1 ($F[23,736] = 1.51$, *n.s.*, for the experiment-by-picture interaction) despite the fact that halving the image size doubled the physical/retinal spatial frequencies in each image. This result is consistent with the hypothesis that AM-rate matches depend on object-based spatial frequency, which is independent of image size.

We performed the same analyses we did in Experiment 1 to investigate how AM-rate matches depend on spatial frequency and perceived clutter. If the relationship between spatial frequency and AM-rate match depends on physical/retinal spatial frequency, the correlation should peak at the same spatial frequency in cycles per degree as it did in Experiment 1. Alternatively, because we halved the linear size of each image, if the relationship depends on object-based spatial frequency, the correlation should peak at the doubled spatial frequency in cycles per degree (corresponding to the same object-based spatial frequency). As for the relationship between spatial frequency and perceived clutter, if dense/sparse judgments are based on clutter in an object-based representation, this correlation should also peak at the doubled spatial frequency.

As in Experiment 1, we found that larger contrast energy within specific ranges of spatial frequencies was strongly associated with faster AM-rate matches and with greater perceived clutter (Figure 5). Importantly, the critical spatial-frequency ranges in cycles per degree doubled relative to those in Experiment 1, consistent with the hypothesis that the association between AM rate and spatial frequency and that between perceived clutter and spatial frequency both depend on object-based spatial frequency. Further, we replicated the strong correlation between perceived clutter and AM rate match, $r = 0.80$, $t(22) = 6.35$, $p < .0001$, in an aggregate correlation, which is also significant when computed separately for each participant, $t(13) = 13.14$, $p < .0001$, confirming our inference from Experiment 1 that AM-rate matches are partly based on the perceived clutter of a visual scene. We also replicated

the result that spatial frequency makes a contribution to AM-rate matches over and above perceived clutter; as in Experiment 1, the correlation peak occurred at a higher spatial frequency for perceived density ($M=2.20$ c/d, $SD=2.92$) than for AM-rate matches ($M=1.54$ c/d, $SD=2.82$), $t(11) = 2.84$ $p = .01$ (compare Figures 3 and 5). The replication of Experiment 1 with respect to doubled spatial frequency in cycles per degree suggests that both AM-rate matches and perceived clutter depend on object-based spatial frequency.

Discussion

Our results have revealed a novel relationship between auditory rhythm and visual clutter. Participants matched images to faster or slower AM rates based on the strength of a specific range of object-based spatial frequencies that indicated the numerosity of object boundaries and contours associated with perceived clutter.

Interestingly, the spatial frequencies most closely associated with faster AM-rate matches and those most closely associated with greater perceived clutter are similar but significantly different. This shows that specific spatial frequency components contribute to AM-rate matches over and above their contributions via generating the explicit perception of visual clutter. It is possible that the object-boundary and contour information conveyed by the critical spatial frequency range (Figure 4) implicitly drives faster AM-rate matches. Alternatively, it is possible that the contrast energy in this spatial frequency range drives faster AM-rate matches by generating other characteristic perceptual experiences, such as dynamism and implied loudness. Investigating these possibilities would be an interesting avenue for a future investigation.

An important difference between the current results and those of Guzman-Martinez et al. (2012) is that, whereas their AM-rate matches depended on physical spatial frequency, ours depended on object-based spatial frequency. Because Gabor patches might simulate visual perception of corrugated surfaces, Guzman-Martinez et al. (2012) suggested that the crossmodal association between physical spatial frequency and AM rate might derive from multisensory experience of manually exploring textured surfaces. That is, if we assume that the speed of manual exploration is relatively constant, the density of surface corrugation conveyed by physical spatial frequency is positively correlated with the AM rate of the sound generated by gliding a hand over the surface. Of course, one cannot glide a hand over an entire natural scene. It is possible that the dual nature of visual information, conveyed as both a texture and a collection of visual objects, may give rise to a texture-based association with auditory AM rate in terms of physical spatial frequency and an object-based association with auditory AM rate in terms of object-based spatial frequency. Another possibility is that because abstraction of perceptual relationships is common in perception and cognition (e.g. Baliki, Geha, Apkarian, 2009; Barsalou, 1999; Piazza, Pinel, Le Bihan, & Dehaene, 2007; Walsh, 2003), it is conceivable that the crossmodal association between surface corrugation (reflected in physical spatial frequency) and AM rate developed through the multisensory experience of manual exploration, might extend to a more abstract crossmodal association between clutter (reflected in object-based spatial frequency) and AM rate. Future research is necessary to evaluate these and other possibilities.

Overall, our results suggest that the auditory-visual association between rhythm and clutter is a fundamental synesthetic association akin to those between loudness and brightness and between pitch and lightness/elevation/size. Our results also suggest that visual clutter is a scene feature that people spontaneously associate with auditory rhythm because our participants were not instructed to use any specific strategy to generate the auditory rhythm that matched each visual image.

Importantly, previous research investigating behavioral effects of auditory-visual associations suggests that the association between auditory rhythm and visual clutter may have behavioral consequences. Firstly, it has been shown that sounds can bias the perception of associated visual features. For example, it was recently shown that hearing laughter makes a single happy face appear happier and makes a sad face in a crowd appear sadder (Sherman et al., 2012). Additionally, hearing a /wee/ sound, typically produced by horizontally stretching the mouth, makes a flat ellipse appear flatter (Sweeney et al., 2012). Likewise, an auditory rhythm may bias the overall perceived visual clutter of a scene. For example, it is possible that listening to a fast (or slow) rhythm might increase (or decrease) the perceived visual clutter of a scene.

Another potential behavioral consequence of the correspondence between rhythm and clutter is based on the finding that sounds can guide attention and eye movements to associated visual objects or features. For example, hearing a “meow” sound guides eye movements towards and facilitates the detection of a target cat in visual search (Iordanescu et al., 2011). Additionally, hearing an AM sound guides attention to a Gabor patch carrying the associated spatial frequency (Guzman-Martinez et al., 2012). It is thus possible that a faster auditory rhythm might guide attention and/or eye movements to more cluttered regions in a visual scene. In considering these potential behavioral consequences, it will be important to determine whether the association generalizes to other ways of presenting auditory rhythm such as music. We used amplitude-modulated white noise because amplitude modulation is the simplest way to convey auditory rhythm and a white-noise carrier contains a broad range of auditory frequencies so that our results are not idiosyncratic to any specific auditory frequency.

In summary, we have demonstrated a novel perceptual association between auditory rhythm and visual clutter conveyed by a specific range of object-based spatial frequencies. This crossmodal association may allow the use of auditory rhythms to modulate the impression of visual clutter as well as to guide attention and eye movements to more cluttered regions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

- Baliki MN, Geha PY, Apkarian AV. Parsing pain perception between nociceptive representation and magnitude estimation. *Journal of Neurophysiology*. 2009; 101:875–887. [PubMed: 19073802]
- Barsalou LW. Perceptual symbol systems. *Behavioral and Brain Sciences*. 1999; 22:577–660. [PubMed: 11301525]
- Bhatara A, Tirovolas AK, Duan LM, Levy B, Levitin DJ. Perception of Emotional Expression in Musical Performance. *Journal of Experimental Psychology: Human Perception and Performance*. 2011; 37(3):921–934.
- Bernstein IH, Edelman BA. Effects of some variations in auditory input upon visual choice reaction time. *Journal of Experimental Psychology*. 1971; 87(2):241–247. [PubMed: 5542226]
- De Valois RL, Albrecht DG, Thorell LG. Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*. 1982; 22:545–559. [PubMed: 7112954]
- Evans KK, Treisman A. Natural cross-modal mappings between visual and auditory features. *Journal of Vision*. 2010; 10(1):6, 1–12.
- Geisler WS, Albrecht DG. Visual cortex neurons in monkeys and cats: detection, discrimination, and identification. *Visual Neuroscience*. 1997; 14:897–919. [PubMed: 9364727]
- Guzman-Martinez E, Ortega L, Grabowecky M, Mossbridge J, Suzuki S. Interactive Coding of Visual Spatial Frequency and Auditory Amplitude-Modulation Rate. *Current Biology*. 2012; 22:383–388. [PubMed: 22326023]

- Iordanescu L, Grabowecky M, Franconeri S, Theeuwes J, Suzuki S. Characteristic sounds make you look at target objects more quickly. *Attention, Perception, & Psychophysics*. 2011
- Juslin PN, Laukka P. Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code? *Psychological Bulletin*. 2003; 129(5):770–814. [PubMed: 12956543]
- Landy, MS.; Graham, N. Visual perception of texture. In: Chalupa, LM.; Werner, JS., editors. *The Visual Neurosciences*. Cambridge, MA: MIT Press; 2004. p. 1106–1118.
- Liang L, Lu T, Wang X. Neural representations of sinusoidal amplitude and frequency modulations in the primary auditory cortex of awake primates. *Journal of Neurophysiology*. 2002; 87:2237–2261. [PubMed: 11976364]
- Marks LE. On cross-modal similarity: auditory-visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception & Performance*. 1987; 13(3):384–394. [PubMed: 2958587]
- Mossbridge J, Grabowecky M, Suzuki S. Changes in auditory frequency guide visual-spatial attention. *Cognition*. 2011; 121(1):133–139. [PubMed: 21741633]
- Olshausen, BA.; Field, JD. What is the other 85 percent of V1 doing?. In: Leo van Hemmen, J.; Sejnowski, TJ., editors. *23 Problems in Systems Neuroscience*. New York: Oxford University Press; 2006. p. 182–221.
- Parish & Sperling. Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination. *Vision Research*. 1991; 31(7/8):1399–1415. [PubMed: 1891827]
- Piazza M, Pinel P, Le Bihan D, Dehaene S. A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron*. 2007; 53:293–305. [PubMed: 17224409]
- Ramachandran VS, Hubbard EM. Synaesthesia – a window into perception, thought and language. *Journal of Consciousness Studies*. 2001; 8(12):3–34.
- Scherer KR. Vocal affect expression: A review and a model for future research. *Psychological Bulletin*. 1986; 99:143–165. [PubMed: 3515381]
- Schreiner CE, Urbas JV. Representation of amplitude modulation in the auditory cortex of the cat. I. The anterior auditory field (AAF). *Hear. Res*. 1986; 21:227–241. [PubMed: 3013823]
- Schreiner CE, Urbas JV. Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields. *Hear. Res*. 1988; 32:49–63. [PubMed: 3350774]
- Schyns PG, Oliva A. From blobs to boundary edges: Evidence for time and spatial scale dependent scene recognition. *Psychological Science*. 1994; 5:195–200.
- Shintel H, Nusbaum HC. The sound of motion in spoken language: Visual information conveyed by acoustic properties of speech. *Cognition*. 2007; 105:681–690. [PubMed: 17196190]
- Sherman A, Sweeny TD, Grabowecky M, Suzuki S. Laughter exaggerates happy and sad faces depending on visual context. *Psychonomic Bulletin and Review*. 2012; 19(2):163–169. [PubMed: 22215467]
- Shulman GL, Sullivan MA, Gish K, Sakoda WJ. The role of spatial-frequency channels in the perception of local and global structure. *Perception*. 1986; 15(3):259–273. [PubMed: 3797200]
- Sowden PT, Schyns PG. Channel surfing in the visual brain. *Trends in Cognitive Sciences*. 2006; 10(12):538–545. [PubMed: 17071128]
- Sweeny T, Guzman-Martinez E, Ortega L, Grabowecky M, Suzuki S. Sounds exaggerate visual shape. *Cognition*. 2012; 124:194–200. [PubMed: 22633004]
- Walsh V. A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*. 2003; 7:483–488. [PubMed: 14585444]

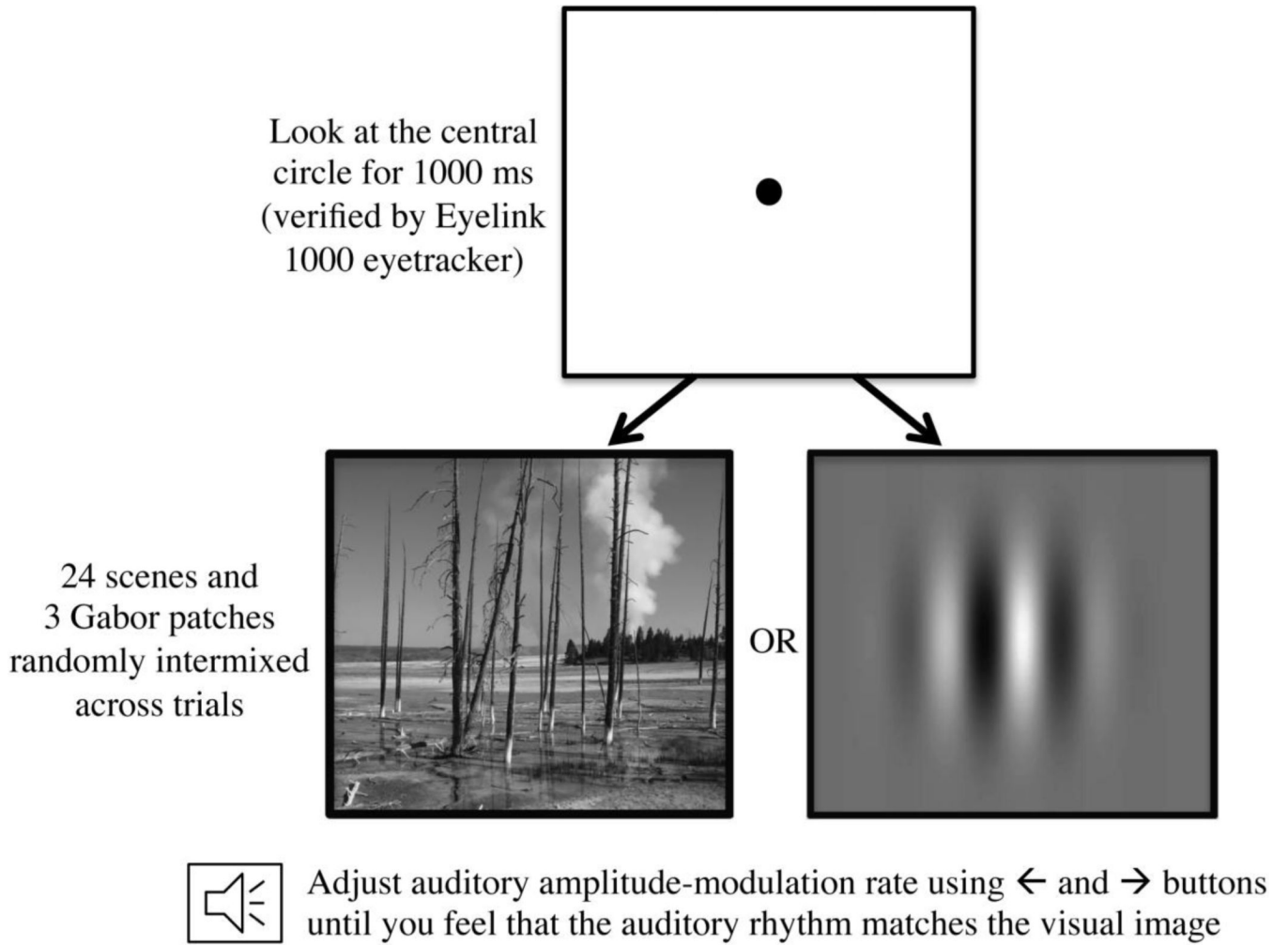


Figure 1.

Trial sequence. Following 1000 ms of central fixation, participants saw a grayscale photograph (27.7° by 20.6° in visual angle) of an outdoor nature, outdoor urban, or indoor scene, or a Gabor patch of one of three spatial frequencies. 1000 ms after the onset of the visual image, participants heard an amplitude-modulated white noise (62 dB SPL) through Sennheiser HD 256 headphones (10 Hz – 20,000 Hz frequency response). While looking at the image, they adjusted the amplitude-modulation (AM) rate of the sound using the arrow keys (increasing or decreasing the AM rate in 1 Hz increments) over the range of 1 Hz to 12 Hz. To avoid an anchoring effect, the initial AM rate was randomly set to 4 Hz, 6 Hz, or 8 Hz on each trial. When participants felt that the auditory rhythm matched the visual image, they pressed a button to register the response. The image then disappeared and the next trial began after a 1000-ms blank screen. A similar crossmodal matching procedure was used in Guzman-Martinez et al (2012).

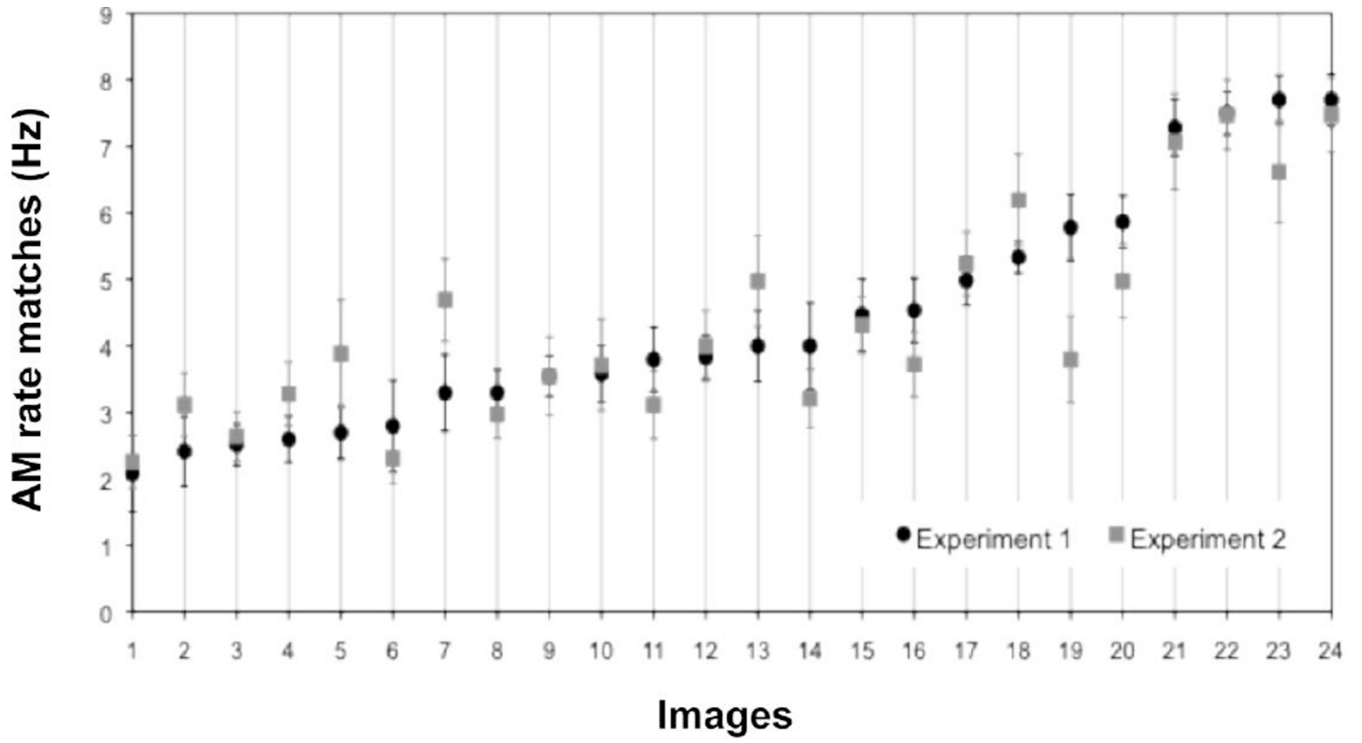


Figure 2. Amplitude-modulation (AM) rate matches to 24 images, ordered from those matched to the slowest AM rate to those matched to the fastest AM rate based on the result from Experiment 1 (black circles). Note that the AM-rate matches are similar in Experiment 2 (gray squares).

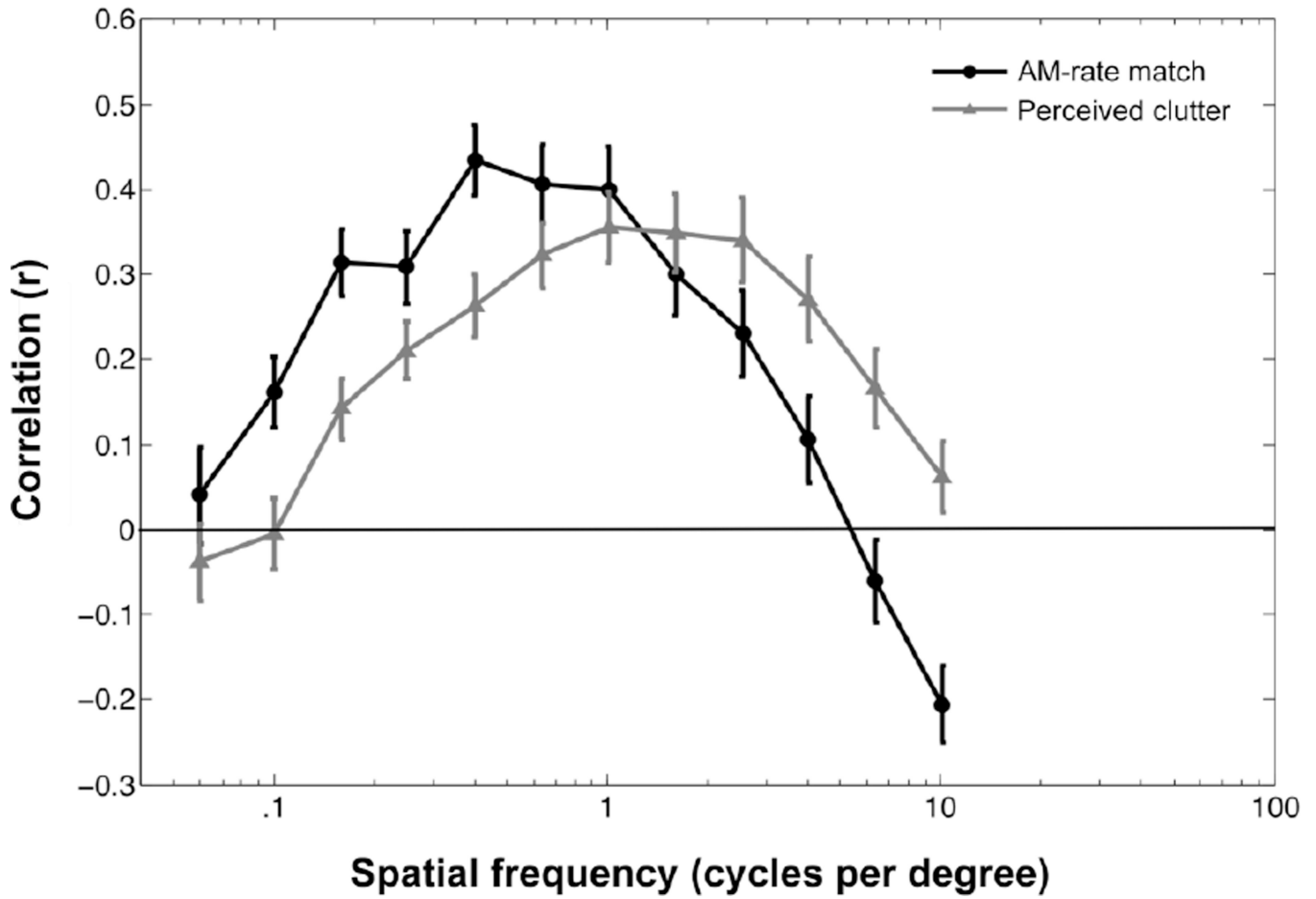


Figure 3. Contributions of the contrast energy in each spatial-frequency bin to amplitude-modulation-rate (AM-rate) match and perceived clutter in Experiment 1. Each point represents the correlation (across images) between the contrast energy in each spatial frequency bin and the matched AM rate (black symbols) or the clutter rating (gray symbols). The error bars represent ± 1 SEM across participants.

Original image

Band-pass filtered (0.3-1.25 c/d)

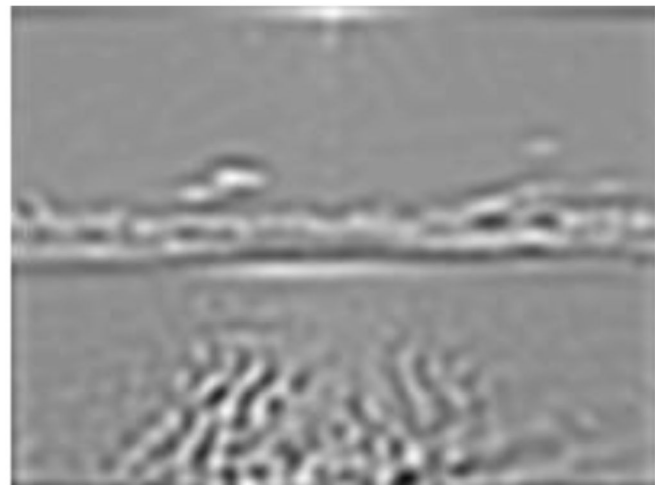
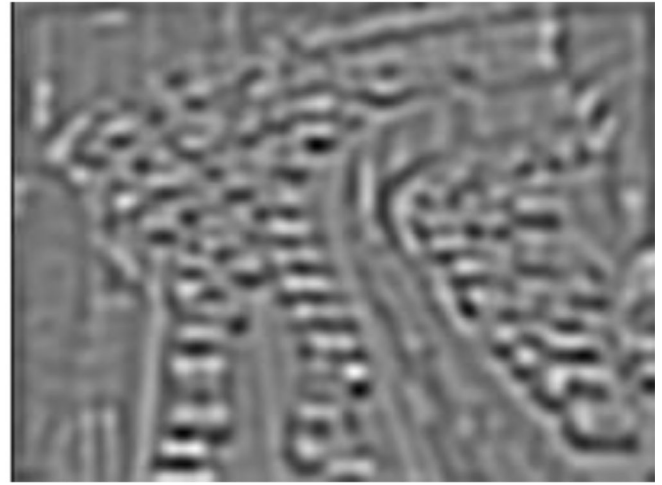


Figure 4.

Two examples of images band-pass filtered at the critical mid-range spatial frequencies strongly associated with faster amplitude-modulation-rate (AM-rate) matches. Comparison of the original and filtered versions of the two example images shows that larger contrast energy in this mid-range spatial frequencies (top images) reflects more object boundaries and contours.

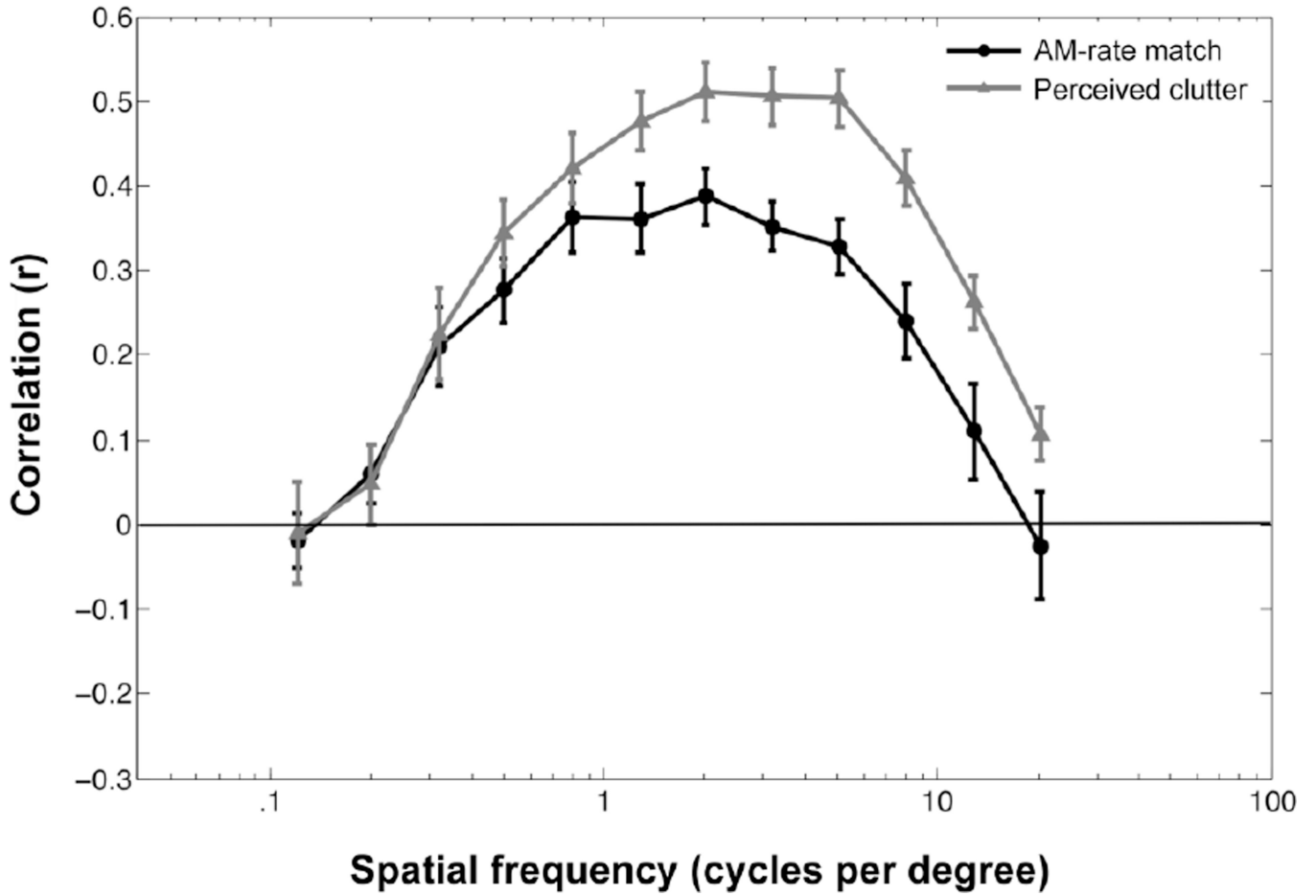


Figure 5. Contributions of the contrast energy in each spatial-frequency bin to amplitude-modulation-rate (AM-rate) match and perceived clutter in Experiment 2. Each point represents the correlation (across images) between the contrast energy in each spatial frequency bin and the matched AM rate (black symbols) or the clutter rating (gray symbols). The error bars represent ± 1 SEM across participants.

Table 1

AM-rate matches to single visual spatial frequencies presented as Gabor patches in Experiment 1 as compared with Guzman-Martinez et al.'s (2012) results for the same spatial frequencies. Spatial frequencies (SF) are indicated as physical spatial frequencies in cycles per centimeter to facilitate comparison with the results presented in Guzman-Martinez et al. (2012).

SF (cycles/cm)	Experiment 1 AM- rate matches (Hz)	Guzman-Martinez et al. AM-rate matches (Hz)
0.50	2.91 (SD=1.45)	2.91 (SD=1.33)
2.20	5.80 (SD=2.04)	6.40 (SD=1.11)
4.50	6.93 (SD=2.61)	9.37 (SD=2.01)

Table 2

For each spatial frequency (SF) bin, the lower and upper boundaries are indicated in both cycles per degree and cycles per pixel. The SF bins are logarithmically spaced.

SF bin	Cycles per degree	Cycles per pixel
1	0.0503 – 0.0796	0.0012 – 0.0019
2	0.0796 – 0.1257	0.0019 – 0.0030
3	0.1257 – 0.2011	0.0030 – 0.0048
4	0.2011 – 0.3184	0.0048 – 0.0076
5	0.3184 – 0.5069	0.0076 – 0.0121
6	0.5069 – 0.8001	0.0121 – 0.0191
7	0.8001 – 1.2693	0.0191 – 0.0303
8	1.2693 – 2.0150	0.0303 – 0.0481
9	2.0150 – 3.2005	0.0481 – 0.0764
10	3.2005 – 5.0814	0.0764 – 0.1213
11	5.0814 – 8.0640	0.1213 – 0.1925
12	8.0640 – 12.8019	0.1925 – 0.3056