Selective attention to the parameters of a physically informed sonic model

Stephen Lakatos

Department of Psychology, Washington State University, Vancouver, Washington 98686
lakatos@vancouver.wsu.edu

Perry C. Cook

Departments of Computer Science and Music, Princeton University, Princeton, New Jersey 08544
prc@cs.princeton.edu

Gary P. Scavone

Center for Computer Research in Music and Acoustics (CCRMA), Stanford University, Stanford, California, 94305
gary@ccrma.stanford.edu

Abstract: Two experiments tested listeners' ability to attend selectively to the properties of a physical model comprising collisions between multiple independent sound-producing objects. A probe signal paradigm measured attention to two properties – resonant frequency and number of colliding objects. Listeners completed a baseline task measuring absolute sensitivity at each stimulus against a background noise. Subsequently, stimuli served as both cues and targets; cue validity was probabilistic. When cue and target were generated by the same object (Experiment 1), greater detectability occurred with valid cues for both resonant frequency and object number, implying the presence of attentional mechanisms for these properties. When cue and target were generated by different objects (Experiment 2), selective attention persisted for object number but not for resonant frequency.

PACS numbers: 43.66.J, 43.60.G, 43.75W

1. Introduction

There is growing evidence that the auditory system can parse and represent the physical properties of sound-generating sources. For example, Warren and Verbrugge (1984) discovered that listeners can distinguish between "breaking" and "bouncing" sounds of glass bottles by attending to their higher-order temporal properties. Lakatos, McAdams, and Caussé (1997) used a crossmodal matching task to measure listeners' ability to detect differences in the width-height ratios of steel and wooden bars and found that listeners are able to distinguish such dimensions by parsing the principal vibrational modes of the bars.

If physical properties of sound sources can be presented auditorially, can we attend selectively to them? This is a difficult question to address with natural sources, since it is often cumbersome to isolate individual physical attributes. Recently developed tools that can permit such manipulations, however, are physical models of sound sources implemented digitally. A number of such models is now available, ranging from exhaustive syntheses based on finite difference solutions of sound sources to computationally efficient, heuristics-based techniques that model the most perceptually salient acoustic properties of a source. To measure selective attention to source properties, we used an algorithm designed by Cook (1997) that models random collisions between particles using parametric stochastic synthesis.

Studies of auditory attention using variants of the auditory probe signal paradigm (Greenberg and Larkin, 1968) have consistently demonstrated selective attention to frequency, spatial location, and intensity (see Scharf, 1998). Recent neuropsychological findings support the notion of sensory gating mechanisms involved in attending to these attributes (e.g., Alcaini *et al.*, 1994). In the current work, we used the probe signal paradigm to examine selective attention not to frequency or spatial location, but to the acoustic properties of a physical model of colliding objects. Using a two-interval forced choice method, we first determined signal levels that produced a uniform detectability across all stimuli for individual subjects. We then measured detectability of the same stimuli under conditions in which one signal was expected (attended) and the other was not (unattended). Attention was manipulated by cueing subjects before each trial with a weak suprathreshold signal of the kind that was likely to be presented. Greater detectability for attended signals would indicate an effect of selective attention for acoustic properties of the cue.

2. Physically Informed Sonic Modeling (PhISM)

The goal of physically informed sonic modeling (PhISM) is to couple physical simulations to efficient synthesis techniques (Cook, 1997). PhISEM (physically informed stochastic event modeling) is a PhISM algorithm based on particle models. In PhISEM, models of particles in containers were solved numerically using the basic Newtonian equations of motion. Simulations with varying numbers of particles and damping (loss of energy when particles collide with each other or the container) were run and statistics were collected about the likelihood of a sound-producing collision, the overall decay in sound energy, etc. Sound is produced only by particles hitting the container shell, because collisions between particles do not couple efficiently to the radiated sound.

The resulting PhISEM synthesis algorithm reduces the behavior of particle systems to a statistical process in which parameters relate directly to the parameters collected in the direct simulations. System energy, which represents the total kinetic energy in the system, decays exponentially. This exponential decay is rapid for systems with high damping. There is a Poisson probability of sound-producing collisions with a high waiting time for few objects and a low waiting time for many objects. This model approaches the ideal for larger numbers of particles. Sound-producing events are modeled as a short exponentially decaying of white noise, and the system resonances are modeled using biquad resonance filters. Even though the original models studied were particles within a sphere (beans within the gourd of a virtual maraca), PhISEM extends well to other systems with multiple independent sound-producing objects. From a psychoacoustic standpoint, the PhISEM model is appealing because it permits control over several physical parameters including the number of colliding objects, the damping properties, or the resonant frequency of the gourd (or chimes) themselves. We selected two parameters - the number of objects and the resonant frequency - to determine whether listeners could attend to them.

The bamboo wind chime and guiro models are different in two important ways. In the guiro, the resonance is fixed, modeling the gourd, whereas in the wind chimes, the resonance parameter is an average center resonance of a distribution of chime frequencies. Each time there is a collision in the chimes, a new frequency is allocated randomly +/- 20 percent around the target center resonance. The number of objects parameter expresses the (statistical) number of bamboo cylinders in the wind chimes. The guiro number of objects parameter applies to how many serrations are caught by the stick as the guiro is scraped.

3. Experiment 1

3.1 Method

Subjects were ten undergraduates between the ages of 18 and 40 recruited from Washington State University and compensated for their efforts. None reported any hearing problems.

Four stimuli were generated with the PhISEM model of the bamboo chimes by crossing two levels of resonant frequency - "low" (1.6, 1.7 kHz spectral centroid) and "high" (4.2 kHz centroid) with two levels of object number - "few" (4-6 objects) and "many" (28-32 objects). We selected these values to generate sounds that were clearly able to be discriminated by listeners. A constant shake energy was used to generate all four stimuli. Figures 1a, b, c, and d plot the spectra for the four bamboo chime sounds averaged across the full duration of the sounds. Spectral centroid values are calculated with a formula used by

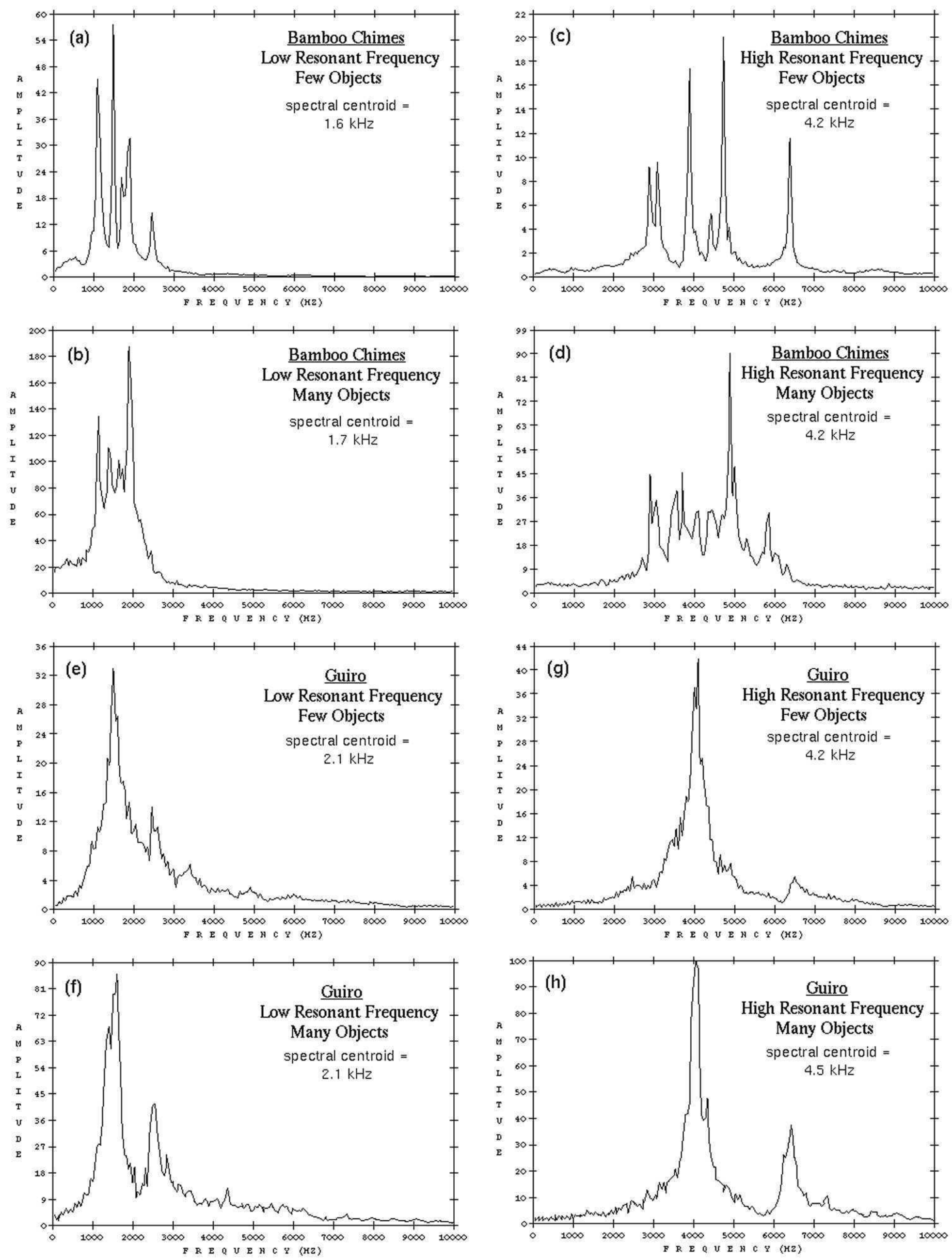

Fig 1. Spectra for the four bamboo chime stimuli in Experiments 1 and 2, and the four guiro stimuli in Experiment 2. Each set of four stimuli was generated by crossing two levels of resonant frequency - "low" (1.6-1.7 kHz centroid for bamboo chimes, 2.1 kHz for guiro) and "high" (4.2 kHz for bamboo chimes; 4.2-4.5 kHz for guiro) - with two levels of object or particle number - "few" (4-6 objects) and "many" (28-32 objects).

Beauchamp (1995). There was little spectral overlap between sounds from sources with low and high resonant frequencies. Those with similar frequencies share several resonance modes, although their peaks shift somewhat depending on the number of colliding objects.

The probe signal paradigm had two parts each lasting about one hour. In the first part, each subject completed a baseline task that measured absolute sensitivity at each of the four sounds against a background noise using an adaptive two-alternative, forced-choice (2AFC) task to ensure that thresholds were independent of criterion and all stimuli were equally detectable. Descending tracks were obtained from each subject using a 2-down, 1-up rule. Signal level was either increased by 1 dB after a single incorrect response or decreased by 1 dB after two consecutive correct responses. Thresholds corresponding to a probability of $0.5^{1/2}$ or 70.7% correct in 2AFC were obtained by averaging across 12 reversal points.

In the second part, separate probe signal tests were run for frequency and object number. For blocks of trials measuring selective attention to frequency, the frequencies of cue-target pairs were randomly selected from among four possible combinations: low-low, high-high, low-high, or high-low. For such blocks, object number was held constant within blocks at either "few" (4-6 objects) or "many" (23-32 objects). Similarly, cue-target pairs for blocks of trials for object number were randomly selected from among the following four sets of values: few-few, many-many, few-many, and many-few. Frequency was held constant within blocks at either a low (1.6, 1.7 kHz) or high (4.2 kHz) value. Measurements obtained from the first part of the experiment were used to calibrate target stimulus levels individually for each subject. Each trial began with a 625-ms cue at a level 15 dB above the subject's threshold for that sound. That cue was followed after 1000-ms by two observation intervals separated by 500 ms, one of which contained a 625-ms target signal (probe) at the level of the subject's threshold: The cue and target were derived from the active "shaking" of the sound object and omitted the exponential decay that followed cessation of the shaking (a 50-ms artificial decay was superimposed at the end of the signal), to prevent subjects from attending simply to differences in decay rates. The subject then responded by indicating which interval contained the signal. Over the session, there was a 75% likelihood that the cue and target shared the same physical property (25% likelihood that the properties differed), and therefore cue validity was probabilistic rather than certain. Greater detectability with valid cues would imply the presence of attentional mechanisms associated with these object properties.

Stimuli were presented with the SigGen/PsychoSig psychoacoustic testing software packages operating in conjunction with Tucker-Davis Technology hardware and running on a Pentium microcomputer. Subjects listened to stimuli over Sennheiser HD265 headphones.

3.2 Results

To compare detection performance across trials containing valid and invalid cues, for each subject we computed the proportions of correct responses for the two types of cues at each resonant frequency and object number. For the analyses of variance, we applied an arcsine transformation to the data to equalize variances. Figure 2 shows the untransformed data in Experiment 1 expressed as the percentage of correct responses for attended and unattended physical properties, shown separately for resonant frequency and object number [Bars for attended stimuli each comprise 720 judgments, whereas those for unattended stimuli comprise 240 judgments.] Detection performance was higher for attended (66.5%) vs. unattended (47.9%) targets generated by bamboo chimes of different resonant frequency [$F(1,9)=22.5$, $p=.001$], indicating that subjects were able to use object number in the detection task. There was no significant difference in detection performance for low (54.5%) vs. high (60.0%) resonant frequency cues pooled across attended and unattended targets [$F(1,9)=2.47$, $p=.15$].

Similarly, when object number varied, detection performance was significantly higher for attended (67.5%) vs. unattended (61.2%) targets [$F(1,9)=6.67$, $p=.03$]. Curiously, detection was better for cues comprising many objects (67.2%) vs. cues with few (61.5%), although the difference misses statistical significance [$F(1,9)=3.8$, $p=.08$]. The reason for this latter performance difference is unclear, although we suspect that because the exponential decay function for sounds generated by many objects was less steep, a relatively longer segment of the signal may have been available at or near listeners' threshold for such sounds.

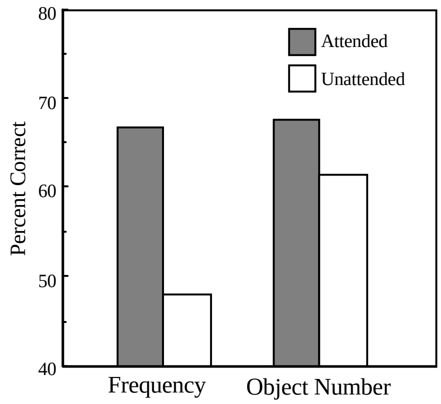

Fig. 2. Listeners' detection performance in Experiment 1.

Results from both the resonant frequency and object number conditions indicate that listeners were able to attend selectively to these properties. The ability to attend to resonant frequency is perhaps not surprising given the considerable evidence for selective attention to pure tone frequency. However, selective attention to the number of colliding objects points to attentional mechanisms more complex than those found for pure tone frequencies, perhaps arising at central levels. In this light, we attempted to increase the complexity of the detection task in Experiment 2 by determining if selective attention to resonant frequency and object number would persist even if cue and target sounds arose from *different* percussive instrument models, which nonetheless shared these two underlying physical parameters.

4. Experiment 2

4.1 Method

The ten subjects from Experiment 1 were also invited to participate in Experiment 2. All agreed to return and were compensated for their participation..

Stimuli were the four bamboo chime sounds used in Experiment 1 and four guiro sounds (see Figures 1e, f, g, and h) generated by crossing two levels of resonant frequency - "low" (2.1 kHz) and "high" (4.2-4.5 kHz) - with two levels of object number - "few" (4-6 objects) and "many" (28-32 objects). Although the spectral centroids of the low frequency guiro sounds were higher than those of the low frequency bamboo chimes - a function of differences in the physical models appropriate to simulating these two percussive instruments - we were confident that they would nonetheless serve as effective cues to resonant frequency. Thresholds for the four guiro sounds were obtained for each subject with the tracking method of Experiment 1. The probe signal paradigm was also retained, except that cue and target on each trial came from different instruments: cue (guiro)-> target (bamboo chimes), or cue (bamboo chimes)->target (guiro). Software/hardware were retained from Experiment 1.

4.2 Results

The untransformed percentage of correct responses for attended and unattended physical properties is shown separately for resonant frequency and object number in Figure 3. With different percussive instruments for cue and target, listeners showed no significant difference in detection performance for resonant frequency for attended (58.2%) vs. unattended (55.0%) targets [$F(1,9)=.01$, $p=.92$], contrary to our initial expectations. Perhaps information about resonant frequency cannot be used to direct attentional mechanisms to sounds with different sources and/or highly distinct timbres. Detection of low (56.4%) vs. high (56.8%) resonant frequency cues pooled across attended and unattended targets was not significantly different [$F(1,9) = .233$, $p=.64$]. Listeners were better able to detect attended (61.1%) vs. unattended (51.4%) targets when cued to the number of colliding objects [$F(1,9)=5.19$, $p=.05$], suggesting that the perceptual salience of this parameter transcended the particular physical model (i.e., bamboo chimes or guiro) that it controlled. Detection was no better for cues comprising many objects (54.7%) vs. those comprising few objects (59.7%), [F(1,9)=.41, p=.54].

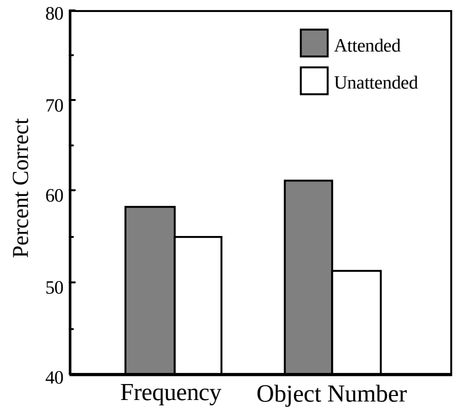

Fig. 3. Listeners' detection performance in Experiment 2.

5. Conclusion

The results reported here point to a fairly remarkable ability of the auditory system to monitor individual acoustical properties of sound sources. Perhaps most impressive was that subjects were still able to attend selectively to the number of objects in Experiment 2, even when the cue and target signals were generated by different objects. Conversely, the absence of selective attention to resonant frequency in this context may point to limitations to frequency-based attention across widely disparate timbres. It remains unclear precisely what features of the proximal waveform may be the most salient "markers" of these distal acoustic properties. However, given the quasi-random nature of the collisions simulated by the PhISM models used here and the ability of listeners to use information about properties in one instrument to attend to those in another, it is clear that the proximal spectral and temporal cues must be highly complex. In future work, we intend to isolate classes of such cues in a number of different physical models to derive a predictive, ecological model of timbre that links the acoustic properties of sounds and their sources to their perceptual correlates.

Acknowledgments

This work was supported by the Air Force Office of Scientific Research (grant #F49620-99-1-0293). We thank Candice Lindsay and Faiza Noor for their help in running participants.

References and links

Alcaini, M., Giard, M-H., Eschalier, J-F., and Pernier, J. (**1994**). "Selective auditory attention effects in tonotopically organized cortical areas: A topographic ERP study," Human Brain Mapping, **2**, 159-169.

Beauchamp, J. W., and Horner, A. (**1995**) "Wavetable interpolation synthesis based on time-variant spectral analysis of musical sounds," Audio Engineering Society Preprint No. 3960.

Cook, P. R. (**1997**). "Physically inspired sonic modeling (PhISM): Synthesis of percussive sounds," Computer Music Journal, **21**, 38-49.

Greenberg, G., and Larkin, W. (**1968**). "Frequency-response characteristic of auditory observers detecting signals of a single frequency in noise: The probe-signal method," Journal of the Acoustical Society of America, **44**, 1513-1523.

Lakatos, S., McAdams, S., and Caussé, R. (**1997**). "Perception of auditory source characteristics: Simple geometric form," Perception & Psychophysics, **59**, 1180-1190.

Scharf, B. (**1998**). "Auditory attention: The psychoacoustical approach," in *Attention*, edited by H. Pashler et al. (Psychology Press, Hove, U.K.).

Warren, W., and Verbrugge, R. (**1984**). "Auditory perception of breaking and bouncing events: A case study in ecological acoustics," Journal of Experimental Psychology: Human Perception & Performance, **10**, 704-712.

