

Perception of Virtual Auditory Shapes

Ari J. Hollander

Email: ari@hitl.washington.edu

Thomas A. Furness III

Email: tfurness@hitl.washington.edu

Human Interface Technology Laboratory

University of Washington, FJ-15

Seattle, WA 98195

Abstract

Virtual environments may help us to understand and eventually extend the domain of auditory perception. The authors performed two experiments which verified the ability of subjects to recognize geometric shapes and alphanumeric characters presented by sequential excitation of elements in a “virtual speaker array.” Performance on the identification task was significantly above chance (33% correct from a field of 10 choices and 42% from a field of 6 choices). The authors noted a highly significant inverse linear dependence of performance on the number of pixels in a pattern. They interpreted this dependency as an indication that the task was cognitively loaded. Analyses of common misidentifications support the hypothesis that a major component of the pattern recognition process derives from perception of the overall motion. These experiments have shown that virtual reality is a viable substitute for the complex apparatus that future work would otherwise require.

1 Introduction and Impetus

Our world has grown to include abstract data spaces and virtual realities. To solidify and extend our presence in these new environments it may be necessary to engage as many of our senses and in as much detail as possible. Virtual environment technology allows us to optimize reality to suit our perceptual systems. Perhaps we can use this technology to solidify and extend perceptions that are only tenuously available in our natural experience.

Even in our normal, physical environment, sound sources are rarely point-sources. Mechanical sounds like the cough of a furnace igniting, the screech of deforming metal, the hum of a turbine coming up to speed, and the thrumming of the wires with the approach of an electric train, all have an “extent” to them. More natural sounds, such as the sounds of wind, rain, ocean waves, and the creaking of trees under the weight of snow, also have size, and in some sense shape as well. Only rarely do we ascribe much importance to the size and shape of a sound.

The purpose of this investigation¹ was twofold. The first aim was to formally validate the idea of using a virtual sound system for communicating spatial patterns. This was accomplished by repeating with synthetically spatialized sound sources a pattern recognition experiment performed with physical speaker arrays [2]. The second purpose was to implicate auditory shape as a potential domain in which human perception can be expanded. In the course of our investigations we found some evidence that patterns generated on speaker arrays (virtual or otherwise) have cognitive

¹For a much more detailed description of the body of work from which this chapter is drawn, see [1].

limitations. We emerged from this study perhaps with more questions than answers, but armed with a new technique with which to help answer these questions in future work.

2 Form in Auditory Space

In audition the sensory mechanism consists of two unidimensional arrays of unidimensional detectors. These sensors supply a rather weak coupling between our internal three-dimensional model and the external three-dimensional² environment. Many of the details must be inferred, and restored during post processing. With careful control of the input stimuli, and special attention to specific perceptual pathways, we might be able to synthesize unusual perceptions.

Unintentionally, we do this all the time. A stereo system equipped with headphones is an extremely artificial way of perceiving sound. In a natural environment, all the sounds would have source locations (although these locations may at times be ambiguous, or difficult to decipher). With the headphones we have detached the sounds from their sources, and have effectively stuffed them in your ears. If you set your amplifier to “mono” something truly extraordinary occurs: the sound appears to be coming from the inside of your head! How many have experienced this as a naturally occurring phenomenon and lived to tell the tale? This odd illusion is not especially useful (in fact audio engineers go to great pains in order to avoid it). However, there may be other, more useful perceptions available elsewhere.

One might pose the question: is our internal spatial representation for audition the same perceptual space we access with our visual systems? Auerbach and Sperling [4] addressed this as a psychophysical question by assuming separate perceptual spaces and then looking for evidence of a transfer between them. They used two speakers aligned with two lights, and measured the ability of human subjects to determine if pairs of stimuli (two lights, two sounds, or one sound and one light) originated in the same or different positions. They examined the errors in this spatial comparison task and found no extra error involved in going from one modality to the other. They took this as a confirmation of the “common space” hypothesis. From a biological standpoint, EEG data has shown that proximal regions in the superior colliculus in mammals are associated with both auditory and visual directionality [5]. Although this may not be conclusive proof that auditory and visual perceptual space are completely coincident, it is, we believe, sufficient motivation to look for traditionally nonauditory, “vision-like” spatial perceptions in audition.

The subjective auditory experience of a clap of thunder is somehow “larger” than that of a firecracker. While this might seem to be an arbitrary qualitative differentiation based on differences in loudness and frequency spectrum, there is a considerable body of research quantifying the apparent *size*, or “space-filling” properties of sounds. This quantity (unfortunately) termed “tonal volume”³ could be considered a one-dimensional manifestation of auditory shape. It is the first precedent to auditory shapes that we discovered in the literature. The topic of tonal volume experienced a period of relative popularity around the turn of the century [6].

An early study by G. J. Rich [7] uncovered reasonable evidence that tonal volume is a distinct perceptual attribute. One compelling fact was that the just noticeable difference for tonal volume was consistent within subjects, and was different than the subjects’ just noticeable difference for pitch. Unfortunately, due to the difficulty of controlling experimental variables using the equipment of the day (Rich literally used whistles as tone generators, and a pendulum to measure

²Actually, our internal representation might be more accurately described in terms of six perceptual dimensions, as the perception of motion via audition is probably superior to static position sensing. See [3].

³Although this attribute is often called just “volume” in the literature, in order to emphasize its distinction from the common synonym for loudness, we will use the term “tonal volume” in this chapter.

time intervals) the tonal volume studies fell into disfavor for a period of about fifty years.

The more modern studies show that tonal volume is a rather convoluted perceptual dimension. Tonal volume increases with the intensity and duration of a sound, and is inversely proportional to its pitch [6]. There are various hypotheses as to the nature of tonal volume. Some describe it as a cognitive phenomenon stemming from associations formed between types of sounds and their sources. Other theories are more physiological in nature and suggest that tonal volume is correlated with the area of the basilar membrane affected by a sound [8]. Unfortunately, none of the theories of the nature of the phenomenon has been thoroughly explored.

In the earlier research tonal volume is considered a scalar quantity: it has spatial magnitude, but no direction associated with it. It is also treated as an essentially monaural phenomenon. In the binaural domain, Perrott and Buell [8] showed that inter-aural decorrelation has a major effect on perceived horizontal extent. Perrott and Buell were able to show that subjects perceived different horizontal extents when confronted with horizontally offset pairs of speakers playing both frequency-offset tones and decorrelated noise. Over the years, this kind of idea has received a great deal of attention in less academic realms: the entertainment and consumer electronics industries are quite involved in the idea of presenting natural, "spacious" sound. There are a number of commercial products available (for example the "Sound Retrieval System" in some Sony televisions) that are designed to, among other things, "widen" the sound output.

There has been some research in two-dimensional auditory pattern perception as well. The standard technique has been to employ a grid of speakers, and to trace out geometric shapes or alphanumeric characters by sequentially energizing speakers along the outline of the shape. Ruff and Perret [9, 10, 11] performed several such experiments and found that their experimental subjects identified the forms at a rate significantly above chance (in one case 36% correct when selecting from 24 possibilities). These studies had the disadvantage of employing sinusoidal stimuli, which are quite difficult to localize off of the horizontal plane. Lakatos' study [2] used a harmonic complex consisting of a 1000 Hz fundamental with a total of 12 partials. This stimulus is better than a pure sinusoid for vertical localization. His subjects performed at an accuracy level of 60-90% from a field of 10 possible choices.

There is substantial evidence that humans can perceive spatial patterns and forms in the auditory domain. Our current research was inspired by some work by Brian Karr and Tom Furness at the Human Interface Technology Lab at the University of Washington. In efforts to make audible pictures, Karr and Furness used virtual sound technology to build an "audio raster" which used a sound source much in the same way that an electron gun is used in a CRT. The sound source was swept back and forth in space as the intensity of the sound was modulated, creating an image. Although this system appeared to be functional (people were able to identify letters), it was never formally tested or developed beyond its initial stage.

3 Experiments

We conducted two experiments with two different sets of shapes to test the perception of virtual auditory shapes. These experiments were shape identification tasks wherein subjects were presented with an auditory pattern over headphones and then requested to choose the picture that this pattern most closely matched.

4 Subjects

Twenty-two subjects were recruited from conferences, Educational Technology classes, and the laboratory staff. These subjects were screened for compatibility with the head-related transfer function (HRTF) used. The best seven were selected for the experiment on the criteria that they experienced a small number of vertical confusions.⁴ Two of the selected subjects were audio engineers, or had extended experience in audio work. By the time of the experiment, all subjects had some experience with virtual spatialized sound sources.

5 Apparatus

For the purpose of conducting the experiment, we built a graphical user interface using Asymmetrix' Toolbook software. Toolbook enables us to easily build and modify graphical interfaces that will accept subject input, record data, control internal devices (such as the ConvolvotronTM), and communicate with external MIDI devices. In the interface for these experiments the subject selects her/his choices from among graphical buttons using a mouse.

The stimulus is presented using a 16-bit digital audio card, SampleCell, by DigiDesign, installed in a Macintosh IIfx. The SampleCell card plays digitized sounds at 44.1k samples/second which are then channeled into a Crystal River ConvolvotronTM card set installed in an Intel 80486DX-based microcomputer. The ConvolvotronTM is configured with the Universal Minimum Phase driver, which consists of an empirically derived HRTF that interpolates between 74 measured spatial positions (12 azimuths, 6 elevations, and the top and bottom poles). Interaural phase information is added after other position-related effects are calculated in order to avoid phase interpolation errors [13]. We used the same HRTF for all subjects.⁵ No head tracking was employed. Consequently, the spatialized sounds were stationary with respect to the subject's head (or rather the subject's headphones), and not the room that the subject occupied.

Subjects listened to the sounds using a pair of Sennheiser HD250B headphones. These headphones were selected because, in addition to their excellent linear frequency response characteristics, they have 16 dB of passive external noise attenuation. As the experiments did not take place in a sound-insulated environment, this characteristic was especially important.

The presentation of instructions, randomization of pattern presentation, and data taking were controlled in software.

6 Stimuli

The stimulus employed was quite similar to that used by Lakatos [2]. It consisted of a 12-partial harmonic complex with a fundamental frequency of 1000 Hz (see Figure 1). The waveform was synthesized using a software package (TurboSynth from DigiDesign) which allows specification of individual frequency components. These waveforms were played back by a Macintosh IIfx using a 16-bit digital audio card, (SampleCell, also from DigiDesign). The characteristics of the stimulus were verified at the point just before they enter the Convolvotron using a LeCroy 9304 digital oscilloscope and performing a power spectrum transformation (see Figure 1).

⁴See [12] for a discussion of vertical confusions with nonindividualized HRTFs.

⁵From the subject SDO in [15].

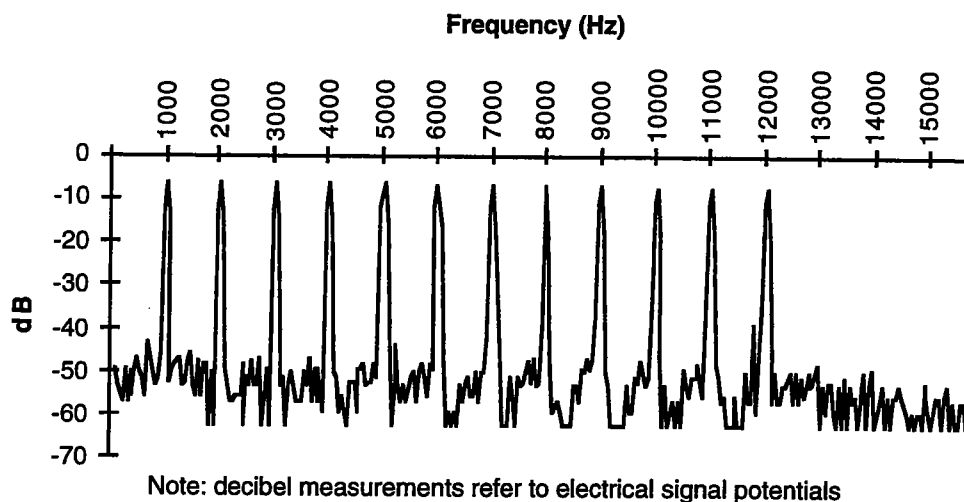


Figure 1: Power spectrum of stimulus.

7 Experiment #1

This experiment was intended to be similar to that of Lakatos [2], only using virtual sound technology. The shape patterns were displayed on a 16 element virtual speaker array. This array was set such that the center of the array appeared at a distance of 6.5 feet from the listener and was oriented perpendicular to the line of sight. Adjacent elements (“speakers”) of the array were spatialized to appear 1 foot apart.

Each virtual speaker, or “pixel,” in a particular shape pattern was, sequentially energized for 60 milliseconds using the stimulus described above. There was a 60 millisecond pause between pixels. The shapes were 10 alphanumeric characters: “3,” “6,” “9,” “C,” “G,” “O,” “P,” “R,” “S,” “U” designed to be analogous to those described in Lakatos [2] (see Figure 2).

8 Procedure

There were 105 trials for each subject: five practice trials, followed by 100 trials during which each shape occurs ten times at random. In each trial, the subject was presented with an auditory shape and was then required to select the best match from screen buttons labeled with alphanumeric characters. The subject was allowed to listen to the shape a maximum of twice before making a selection. For the first 55 trials (including the 5 practice trials) the subject was given feedback. If they selected the correct letter, the button they pressed turned green. If they have made an incorrect choice, the button they selected turned red, and the correct button turned green.

9 Results and Discussion of Experiment #1

A *t*-test comparing correct identification rate to random performance indicates significantly better than chance performance ($T = 12.932, P < 0.001$). An analysis of variance indicated significant main effects for subjects and letters, and a significant interaction between subjects and letters (see Table 1).

In order to better understand the results, we created a confusion matrix with the presented shapes as the column headings and the subject’s choices as the row headings. Next we performed

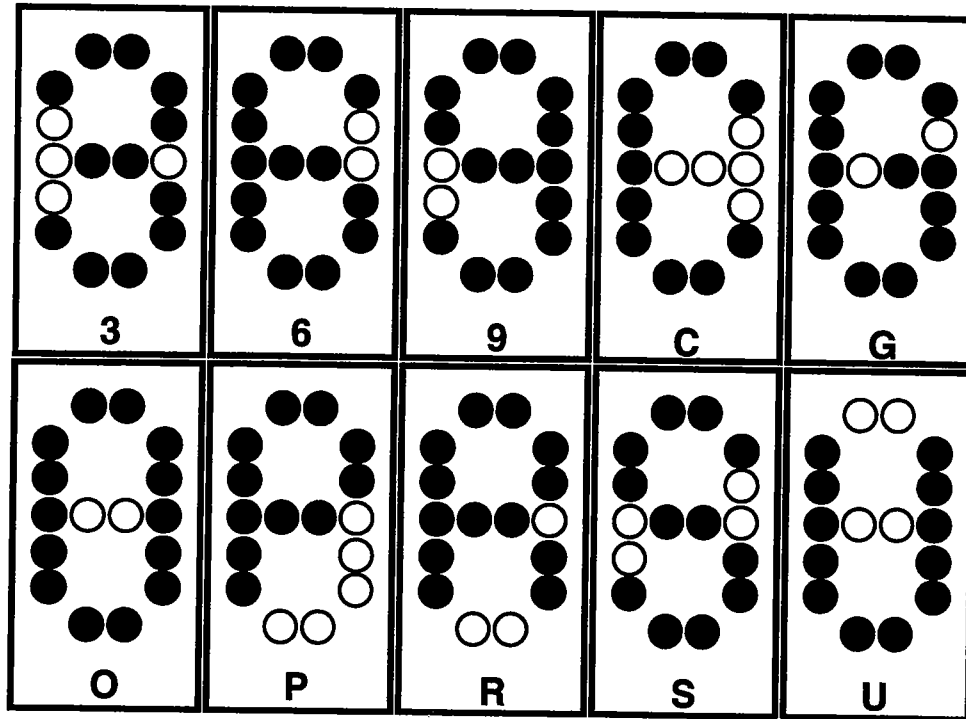


Figure 2: Alphanumeric auditory shape patterns.

a cluster analysis on this matrix and sorted the rows and columns based on clustering (see Figure 3).

This representation brings out some interesting features of the data. As one would hope, the peaks are generally located in the right places; i.e., along the diagonal. The sorting by clusters helps preserve the appearance of the diagonal line in spite of the fact that there are several shapes that were specifically confused. By looking at which shapes fall adjacent to each other, we discover that many confusions are just the ones we would expect: The "O" is the same pattern as the adjacent "C" with three missing pixels. Although the adjacent "S," "6," and "G" are traced in different spatial-temporal patterns, independent of time they are morphologically quite similar.

The diagonal line appears almost "anti-aliased" in places. This is a consequence of the fact that within columns, and to a certain extent within rows, the data is quite smooth. This gives the impression that there is a fair amount of perceptual overlap between shapes that are adjacent in this representation. The one major exception is the region around the correct identification of the letter "C." While the "O" and the "G" are often mistaken for a "C," the reverse is almost never true. This asymmetry probably reflects the fact that the pattern that makes up the letter "C" has only eleven pixels in it, whereas the "O" and "G" have fourteen. This makes the duration of the "C" pattern 360 ms shorter than that of "O" and "G." The difference in duration alone could be the distinguishing factor.

The asymmetry and enhanced performance on the identification of the "C" suggests another analysis: if the patterns with the fewest pixels are the least ambiguous, perhaps there is an overall effect caused by number of pixels in a pattern. We performed an analysis of variance with respect to the number of pixels in each pattern (see Table 2), and discovered that there is a significant effect at the $P < .001$ level. Moreover, post-hoc contrasts show a highly significant *inverse* linear relationship between performance and the number of pixels in a pattern (see Figure 4). One way to interpret this is that the greater complexity of shapes with more pixels confused the subjects. In

Source	Sum-Of-Squares	DF	Mean-Square	F-Ratio	P
Subject	5.480	6	0.913	4.759	0.000
Shape	10.413	9	1.157	6.029	0.000
SubjectXShape	17.977	54	0.333	1.735	0.001
Error	120.900	630	0.192		

Table 1: ANOVA of performance on shape identification task (Experiment 1).

Source	Sum-Of-Squares	DF	Mean-Square	F-Ratio	P
Pixels	8.646	4	2.161	9.612	0.000
Error	250.726	1115	0.225		

Table 2: ANOVA of performance on shape identification task (Experiment 2).

other words, the main challenge to recognizing a pattern was not getting sufficient information, but keeping track of details. This interpretation argues that the pattern recognition task is a difficult *cognitive* exercise, and not a purely perceptual one. Several subjects reported that they would occasionally “lose track” of complicated patterns, and had to resort to guessing.

An alternative explanation which might yield the same result was proposed by Lakatos [2]. He suggested that there may be a bias toward the selection of simpler patterns. This hypothesis is difficult to evaluate without a strong definition of “simplicity.” Two possibilities are depicted in Figure 5.

Overall performance on Lakatos’ [2] experiment (60%–90% correct), with nearly identical patterns and stimuli, was superior to that on our experiment (20%–43% correct). This could be due to the virtual nature of the patterns in our experiment and the fact that we did not use head tracking. Although we believe that these are factors, we would like to show that this does not comprise the majority of the discrepancy in performance. Wenzel, Wightman, and Kistler [12] determined that nonindividualized HRTF’s incur only a minor decrement in acuity of localization. However, they also state that there is a high frequency of front-back, and to a lesser extent, up-down confusions in localization in such situations. Our subjects, however, were selected for their low frequency of vertical confusions.

In order to appropriately compare the two experiments, it is first necessary to emphasize a few differences in methodology other than the virtual-vs-physical aspect. In Lakatos’ experiment, the subjects were given a training session before starting the experiment. In our experiment only enough advance training was supplied to acquaint the subjects with the interface. Instead feedback was supplied for the first half of each test run. We chose this method of training because we wanted to obtain information on how listeners learn the task. Inspection of the change in performance over time in our experiment reveals that there is great variation between subjects. The performance of some subjects increased steadily throughout the half of the experiment in which they had feedback. All but one of these subjects seemed to reach an asymptote in performance level before the end of the feedback (two of these were lower asymptotes, as two subjects started out with a string of correct identifications). Two of the subjects were extremely consistent throughout the experiment run. Since some of the subjects did improve during the experiment. Their initial poor performance is one source of error that is likely a part of the discrepancy between our data and that of Lakatos.

Another methodological difference was that in Lakatos’ experiment, the starting point was depicted, on each pattern choice. This additional cue could greatly affect the distinguishability of several patterns which mainly differed in the direction of tracing, depending on the strategy

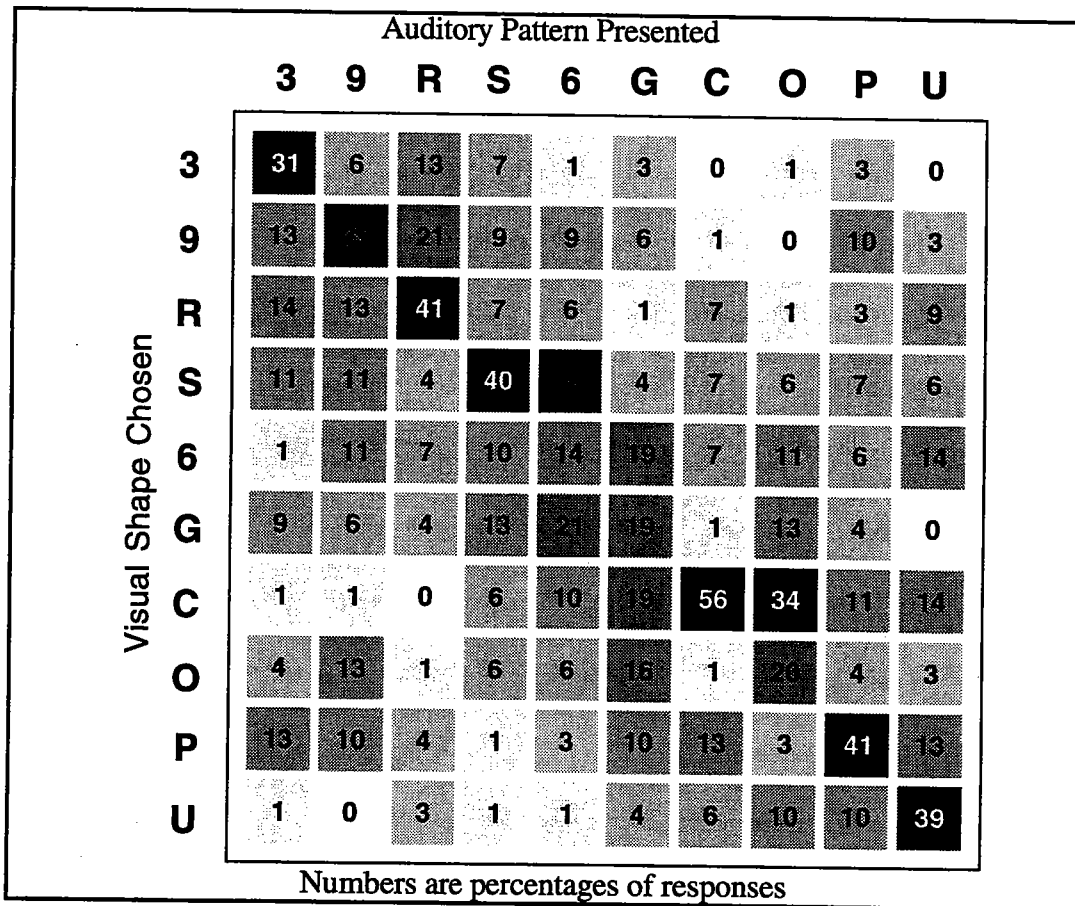


Figure 3: Confusion matrix for Experiment 1.

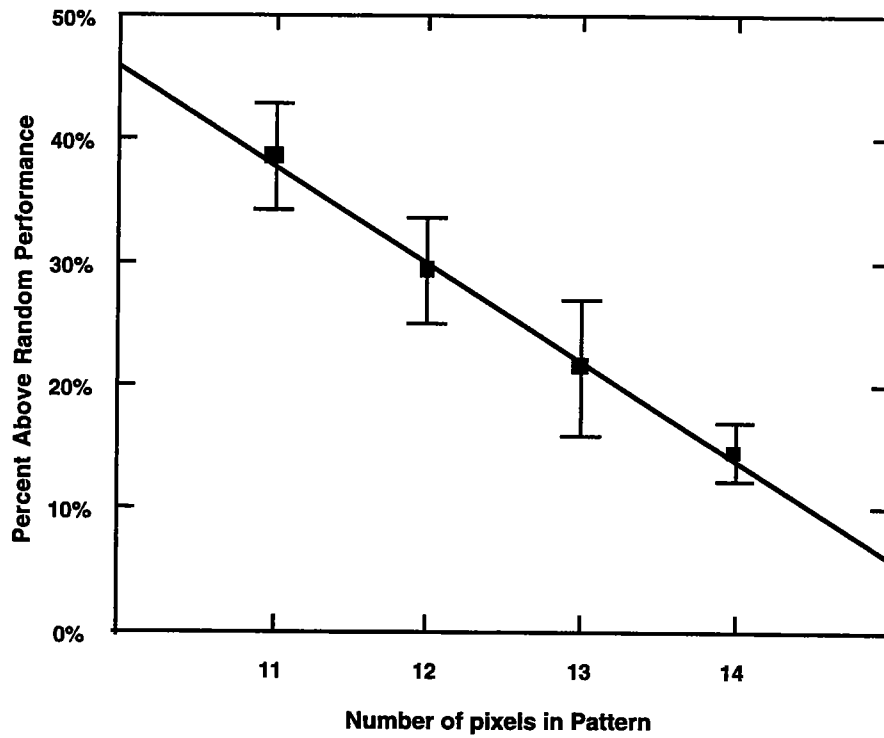


Figure 4: The relationship between identification accuracy and number of pixels.

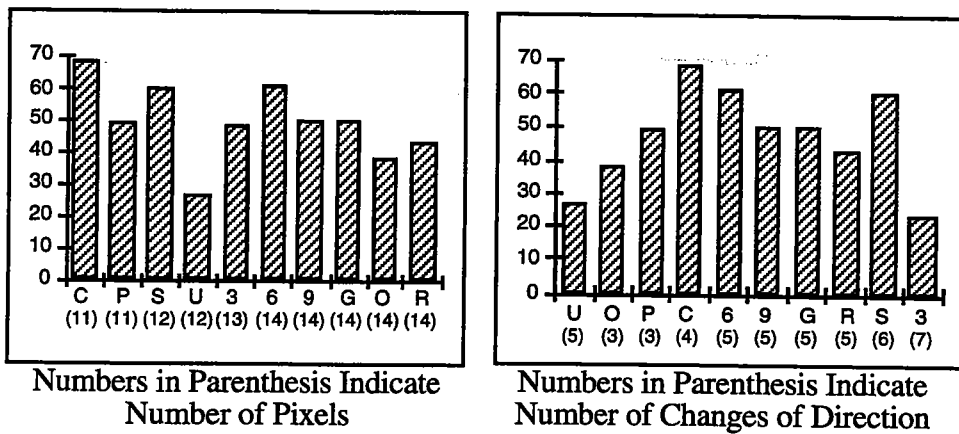


Figure 5: Erroneous selection counts.

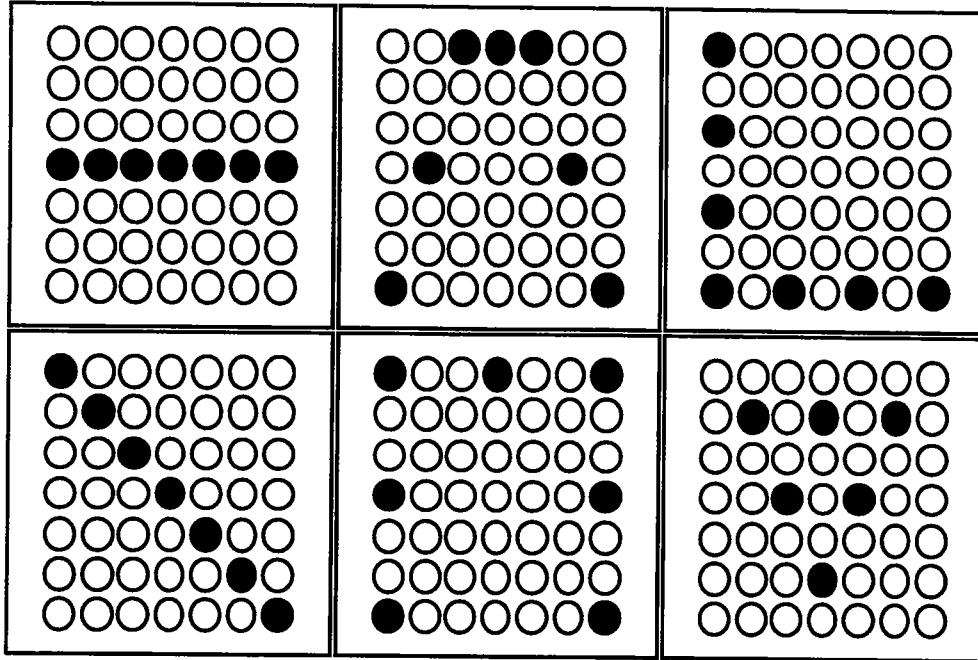


Figure 6: Geometric auditory shape patterns.

used by the subjects. We chose not to supply this cue because we thought that it would bias the subjects' strategies.

To reconcile these differences, we ran one subject who had previously scored quite well on our experiment (43%) through a version of the experiment modified to more closely match Lakatos'. The subject was given a training session in which she had an opportunity to listen to and compare the shapes at will, and to perform a practice run of the experiment. During her practice run she scored 70% (although that was not a balanced set of stimuli where each occurred an equal number of times). Afterward she went through the experiment as she had during the formal trials, only this time the captions on the selection buttons had been modified to include an indication of the starting position and direction of each pattern. She scored 58%, which is close to the performance level in Lakatos' experiment. This subject's dramatic improvement is strong evidence that the majority of the difference in performance between our experiment and Lakatos' was due to methods, and not apparatus.

10 Experiment #2

The patterns in Experiment 1 were quite complex, and thus it was difficult to determine just what aspects of certain shapes made them hard to recognize. Also, we were concerned about the effect of differing numbers of pixels (and hence different pattern durations). We wanted to control these properties a little better.

The patterns in Experiment 2 consisted of six geometric shapes: a horizontal line (from left to right), a diagonal line (from upper left to lower right), a square (with no bottom), a trapezoid (with no bottom), and a triangle. These shapes all have the same number of pixels: seven (see Figure 7). As a consequence of the small number of pixels, the patterns in this experiment are much simpler than those in Experiment 1. The subjects, stimuli, and apparatus were the same as in Experiment 1.

11 Procedure

The procedure in the second experiment is much like that of the first. The only difference is that in this experiment there are only 6 shapes and 6 choices, so there are a total of 66 trials with 6 for practice. Feedback is supplied during only the first 36 trials. Again, each shape occurs 10 times.

12 Experiment #2 Results

The performance on this experiment ranged from 31% to 63% with an overall average of 42%. It should be noted that, as there were fewer choices in experiment 2, this is only 26% above random which is not significantly higher than the 23% above the level of chance performance in experiment 1. Again we performed a cluster analysis of the responses to the shapes, and created a confusion matrix sorted by these clusters (see Figure 8).

In two instances there were highly asymmetric confusions. Shape 2 was mistakenly identified as shape 3 twice as often as it was correctly identified. Even more pronounced was the favoring of shape 5 over shape 6. Further inspection of the confusion rates leads us to believe that the overlap within these pairs is complete. The columns in figure 8 can be thought of as vectors within a shape description space: Shape 1 is 77% like the visual picture 1, 4% like picture 2, 9% like picture 3, and so on. The column profiles of shapes 5 and 6 and also those of shapes 2 and 3 are nearly identical.

The large perceptual overlap between shapes 5 and 6 is not too surprising, as they are morphologically quite similar. The differences between shapes 2 and 3 are more striking. In both cases the favored visual description is the one that most closely follows the overall motion (see Figure 3). This is a strong argument that motion in these patterns is more important than morphology.

An alternative explanation involves the spatial distribution of the pixels in the patterns. The pixels of the two disfavored shapes fall more in the periphery of the grid. This is the region of poorest auditory acuity. The disfavored shapes could therefore be considered the more ambiguous patterns.

The performance on Experiment 2 did not fit into the linear dependency on number of pixels of Experiment 1. This might cloud the interpretation of importance of that factor, but since Experiment 2 was not designed to be especially comparable to Experiment 1, the nature of the patterns themselves is quite different. First, the array used in this Experiment is 3 feet wider than that used in Experiment 1. This means that portions of patterns fall in regions of poorer auditory acuity. And second, since consecutive pattern elements do not necessarily occupy adjacent elements of the array, issues of apparent motion begin to come into play. The largest angular separation between consecutive pixels in this experiment is about 24° (as compared to 11° in Experiment 1). This is within the range of some empirically derived critical values for apparent motion perception with a 120 ms onset interval [16].

Conclusions

We have established that synthetic spatial cues, although they may somewhat degrade performance (as compared with physical speaker arrays), do allow subjects to successfully perform the pattern recognition task. Thus we have validated the use of virtual sound in further study of auditory shape perception. This is quite fortunate, as we have a long way to go before we can make useful auditory shapes. We have identified several issues that need to be studied before we will know if efficient auditory shape perceptions are possible.

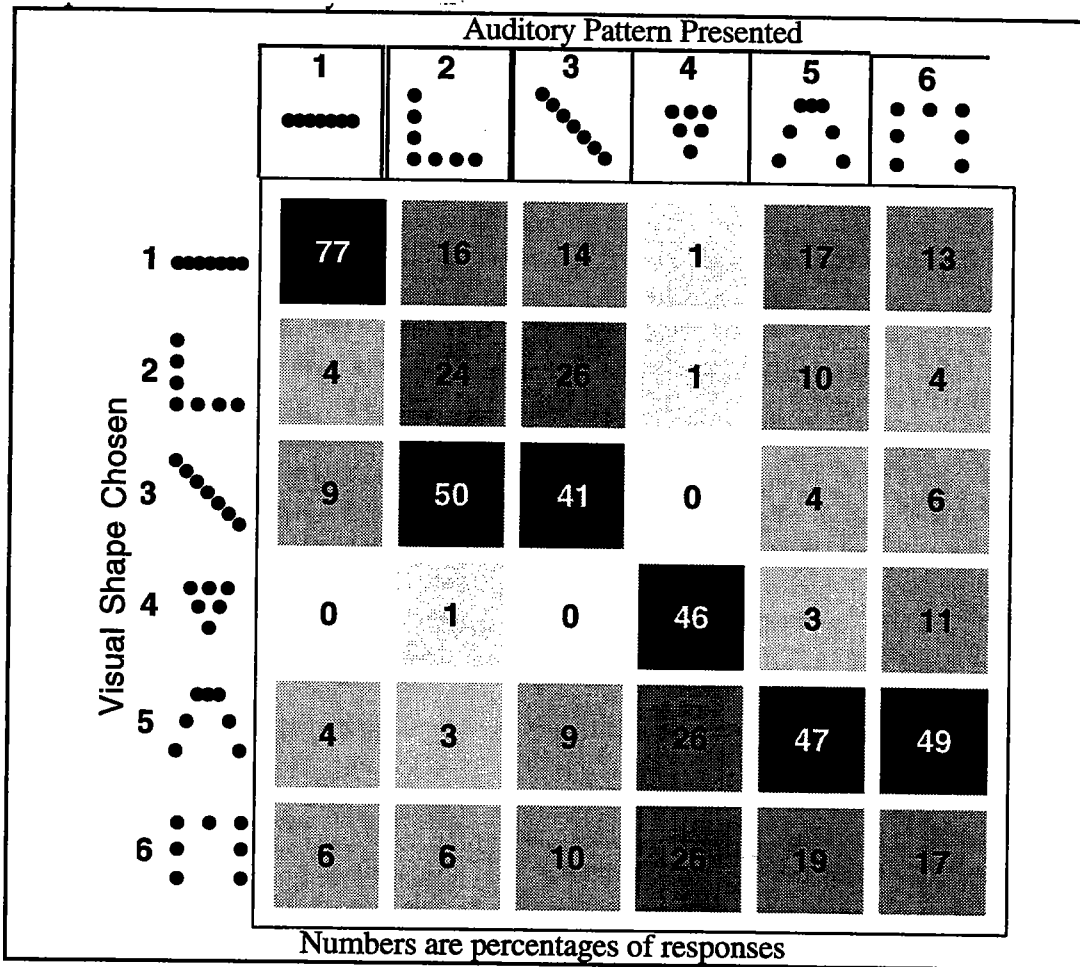


Figure 7: Percentage of responses.

The dependence on the number of pixels in the pattern lends credence to our fear that this particular shape recognition is a largely cognitive process: it is a “perceptualization” and not a true perception. This is not helpful if the intent in conveying shape information via auditory channels is to reduce the overall cognitive load on people in a complex environment. One investigation which is strongly indicated is a measure of the cognitive effort required to perform the shape recognition task.

Our experiments rely heavily on the phenomenon of auditory apparent motion, which is an incomplete representation of sound motion (e.g., it does not include Doppler cues). Further information on the effects of apparent continuity on shape recognition would be quite useful. Similarly, data on the effect of pattern duration and speed would also be useful. Further investigations will be greatly facilitated by the versatility offered by virtual spatialization. The way is also open to examining continuous motion and exploring additional movement cues. For example, the ConvolvotronTM allows for the possibility of including Doppler shifts.

A potentially formidable stumbling block is the inherently time-dependent nature of sequentially presented patterns. Because we need to accumulate information about the spatial patterns over significant time periods, effort may be required in order to integrate this information into a coherent picture. If this is the case, then the sequential presentation paradigm is the wrong approach. Although the auditory system has difficulty with the spatial processing of simultaneous events (as exemplified by the precedence effect) there are some indications that these limitations are not insurmountable. If the concurrent sources are differentiated in some way, then they can retain some of their spatial properties while at the same time manifesting extent [17, 18].

With the use of virtual sound systems we can perform systematic psychophysical experiments with novel types of stimuli. An intriguing idea finds its source in the research of Waugh, Strybel, and Perrott [3]. They showed that humans are as sensitive to velocity with audition as we are with vision. Precise auditory velocity judgments are possible over time intervals as small as 30 ms [19]. To our knowledge there is no research on judgments of concurrent sound-source velocities. If it is determined that humans have facility with such a task, then it may be we could perceive pictures drawn in auditory phase space. Perhaps velocity, not position, is the appropriate primitive for extending auditory experience.

Our experiments, as the first to employ *virtual* audio-spatial patterns, were required to validate the use of synthetic spatial cues in such a process. As such they needed to be substantially grounded in the “rear-view mirror” of previously explored techniques. Now that this has been accomplished, the way is open for more innovative procedures. We have barely begun to exploit the flexibility of virtual sound to generate a wide range of spatial auditory patterns—patterns that would be nearly impossible to create with mechanical apparatus.

13 Acknowledgments

This research was supported by AFOSR grant #F49620-93-1-0339. The authors also wish to thank Dan Shapiro, Brian Karr, Dr. David Wessel, Dr. Woodrow Barfield, and Dr. Robert Duisberg for their assistance, input, and inspiration.

References

- [1] Hollander, A. “An Exploration of Virtual Auditory Shape Perception” Master’s Thesis (1994): <http://www.hitl.washington.edu/projects/audio/hollander/>

- [2] Lakatos, S. "Recognition of Complex Auditory-Apatial Patterns." *Perception* **22** (1993): 363-374.
- [3] Waugh, W., T. Strybel, and D. Perrott. "Perception of Moving Sounds: Velocity Discrimination." *J. Aud. Res.* **19** (1979): 103-110.
- [4] Auerbach, C., and P. Sperling. "A Common Auditory-Visual Space: Evidence for Its Reality." *Percep. & Psychophys.* **16** (1974): 129-135.
- [5] Sparks, D. L. "Neural Cartography: Sensory and Motor Maps in the Superior Colliculus." *Brain, Behavior, and Evol.* **31** (1988): 49-56.
- [6] Perrott, D., Musicant, A., and Schwethelm, B. "The Expanding Image Effect: The Concept of Tonal Volume Revisited." *J. Aud. Res.* **20** (1980): 43-55.
- [7] Rich, G. J. "A Preliminary Study of Tonal Volume." *J. Exper. Psychol.* **1** (1916): 13-22.
- [8] Perrott, D., and T. Buell. "Judgements of Sound Volume: Effects of Signal Duration, Level, and Interaural Characteristics on the Perceived Extensity of Broadband Noise." *J. Acoust. Soc. Am.* **72** (1982): 1413-1417.
- [9] Ruff, R. M. "Audiospatial Integration." *Perceptual & Motor Skills* **60** (1985): 891-902.
- [10] Ruff, R. M., and E. Perret. "Auditory Spatial Pattern Perception Aided by Visual Choices." *Psychol. Res.* **38** (1976): 369-377.
- [11] Ruff, R. M., and E. Perret. "Spatial Mapping of Two-Dimensional Sound Patterns Presented Sequentially." *Perceptual & Motor Skills* **55** (1982): 155-163.
- [12] Wenzel, E., F. Wightman, and D. Kistler. "Localization Using Non-Individualized Head-Related Transfer Functions." *J. Acoust. Soc. Am.* **94** (1993): 111-123.
- [13] Foster, S., W. Chapin, and L. Longley. "Convolutron 3D Audio Interface Minimum Phase HRTF Convolution Technology." Memorandum (1992).
- [14] Wightman, F., and D. Kistler. "Headphone Simulation of Free-Field Listening. I: Stimulus Synthesis." *J. Acoust. Soc. Am.* **85** (1989): 858-867.
- [15] Wightman, F., and D. Kistler. "Headphone Simulation of Free-Field Listening. II: Psychophysical Validation." *J. Acoust. Soc. Am.* **85** (1989): 868-878.
- [16] Lakatos, S. "Temporal Constraints on Apparent Motion in Auditory Space." *Percep. & Psychophys.* **54** (1993): 139-144.
- [17] Perrott, D. "Discrimination of the Spatial Distribution of Concurrently Active Sound Sources: Some Experiments with Stereophonic Arrays." *J. Acoust. Soc. Am.* **76** (1984): 1704-1711.
- [18] Perrott, D., Marlborough, K., and Merrill, P. "Minimum Audible Angle Thresholds Obtained Under Conditions in Which the Precedence Effect is Assumed to Operate." *J. Acoust. Soc. Am.* **85** (1989): 282-288.
- [19] Perrott, D., V. Buck, W. Waugh, and T. Strybel. "Dynamic Auditory Localization: Systematic Replication of the Auditory Velocity Function." *J. Aud. Res.* **19** (1979): 277-285.