Creating Virtual Sound Objects

A. V. Agranovskii and G. E. Evreinov

Sputnikavtomatika State Design Bureau, Gazetnii 51, Rostov-on-Don, 344007 Russia
e-mail: anisi@rs.rnd.runnet.ru
Received February 24, 1997

Abstract—How is an auditory contour pattern perceived? What additional attributes must be adjusted to stabilize the perception of the boundaries of a virtual sound pattern in order to form an auditory perception isomorphic to the visual one? A possibility to correct the perception of subjective patterns through controlling the parameters (the frequency deviation and the duration at the critical points) of virtual sound patterns created by the illusion of the apparent motion of acoustic sources is discussed. © 2000 MAIK “Nauka/Interperiodica”.

Unlike the voluntarily fixable visual attention, for a sound pattern created by the apparent motion of virtual sound sources, it is necessary to efficiently control the parameters that can be used to keep attention of the auditory system continuously focused on the sound sequence developing in time. The attention-control process is quite complicated and insufficiently studied until now.

How is a sound contour pattern perceived? What additional attributes must be adjusted to stabilize the perception of the boundaries of the virtual auditory pattern in order to form an auditory perception isomorphic to the visual one?

We create virtual sound patterns with the help of a vector audio display [1–3]. The device consists of a single-piece acoustic module controlled via a special-purpose interface, which operates four electrodynamical acoustic sources supplied with settings that limit the modified sector to 90°. The sources are placed at the corners of a square with sides 0.46 m in length and at a 0.6-m distance from the auditory channels of the subject, whose head is free (no limitations are placed on the motion) and eyes are closed. The 0.46-m distance was accepted as the base and height of the acoustic plane. The programmable interface creates graphic objects on a monitor screen and then makes a sound pattern of them by moving one or more sound cursors (virtual acoustic sources) formed by controlling the space position of an interference maximum of the sound signal created by these acoustic sources by the rule

\[ A_1 + A_2 + A_3 + A_4 = B = \text{const} \] (1)

or

\[ AXX + A(1 - X)(1 - Y) + AXY + A(1 - X)Y = B, \] (2)

where \( A \) is the signal amplitude, \( X \) and \( Y \) are the current coordinates in the interval [0, 1] with the origin at the lower left point of the image plane, and \( B \) is the subjectively perceived volume of the acoustic source. Therefore, a subjective sound pattern described by the subject can be compared to the real image on the monitor screen. The virtual acoustic plane was additionally simulated by a sheet of paper, on which the tested person reproduced one's subjective impression with a soft-tip pen immediately after observing the presented sound objects while keeping one's eyes closed.

The general view of the part of the interface concerned with creating graphic objects is shown in Fig. 1a. It consists of Windows 95 software modules written in Visual Basic 4.0. The module that creates the graphic objects uses a fixed workspace of 9×9 = 128×128 points. In a sense, it is similar to the matrix of virtual sound sources [4–6], but has an eight times higher spatial resolution and provides the necessary continuous isolationization without applying special interpolation methods to the sound signal between the points. Interpolation is used only for drawing the image.

The toolbar serves to create standard graphic objects, namely, straight lines (a segment or polygonal line), arc, circle, ellipse, rectangle, and arbitrary curve, and to copy and arbitrarily move them within the workspace. The control panel can be used for editing the sound attributes of the graphic objects. An additional panel provides the pointwise editing of the image fragments. This function includes marking the critical points (end points or intermediate ones) that are necessary to control the perception process and adjust an arbitrary number of silent points.

The general view of the sound-control part of the programmable interface is shown in Fig. 1b. The sound frequency can be changed within 100–5000 Hz, though, in our experiments, it was no higher than 800 Hz. The built-in oscillators and attenuators of the acoustic module are controlled with 8-bit analog-to-digital converters to provide a wide choice of sound frequencies and volumes. A relatively wide band of the sound signal was provided by its almost rectangular envelope. The frequency deviation (\( D_1 \) and \( D_2 \)) was controlled as a
function of the image point coordinates \((X, Y)\): in the \(Oy\) axis, independently for each cursor and, in the \(Ox\) axis, simultaneously for all cursors. The initial frequency \(F_0\) was associated with \(X = 0, Y = 128\), i.e., the frequency decreased when the point moved downwards and increased when it moved from the left to the right, so that the lowest frequency is adjusted at the lowest left position of the virtual plane, while the highest frequency is at its upper right point.

Additionally adjustable parameters are the sound duration and the frequency step at the critical points (end or intermediate ones) of each fragment of the sound pattern.
Fragments of the sound pattern created by the related sound cursor can be presented sequentially or in parallel with synchronized start moments, or asynchronously for each cursor: in a single run with possible reversal or in cycles.

The software interface provides a standard set of functions for managing the created files.

Let us study the perception of a fragment of a sound pattern. For better visualization, we use closed contours, in particular, a circumference and a rectangle. Figure 2 shows original graphic objects and their subjective auditory images from graphic reports of the tested persons after a single presentation, without special attributes used to indicate the initial point of the pattern. For the circumference, a sector can be shown around the start point, in which an uncertainty in localizing the trajectory causes significant errors in the perception of the presented sound pattern. The extent of this sector depends on the velocity of the virtual source, radius of the circumference, and signal spectral parameters. Distortions in the perception of the rectangular pattern are minimum near the points where the motion changes its direction, i.e., whose parameters require an additional processing time from the sensor system.

If at a certain time moment, the first (or any other critical) point of the sound pattern appears in the virtual acoustic plane, the new stimulus activates the orientation reaction, which is functionally characterized as the selective attention [8]. The orientation reaction has an excitation threshold and a latent period, which depend on the gradient of the physical parameters of the stimulus and on the initial activation level of the sensor system at the moment when the stimulus is presented. Therefore, the moment of subjective indication of the beginning of the sound pattern not necessarily coincides with the first sound point. The further development of the reaction and perception of the sound sequence substantially depends on the dynamics of the sound stimulus parameters and on such specific parameters of the auditory system as the minimum audible movement angle of the sound source [9]. When the stimulus parameters come to a steady state and stop changing, the attention dissipates. Any change in the stimulus keeps the indicated point in focus and concentrates the attention.

Let us use one sound cursor to represent a trajectory equivalent to a horizontal straight line of 100 points (0.36 m) placed in the central region of the virtual acoustic plane symmetrically relative to its side borders. Set the fundamental frequency on the sound control panel at 305 Hz. Set the frequency deviation along the Ox and Oy axes and the parameters of the additional critical point markers to the inactive state. Then, the subjective perception of the length of the given segment in a single run depends on the time interval within which the sound is emitted by 100 points (sequentially excited virtual sources).
Set the amplitude of the sound cursor at 50 dB. At a 4-ms sound duration at each point, the power of the virtual source (at 305 Hz) may appear to be insufficient to reliably locate the beginning of the segment, which may require the activation of several points. The same effect can be observed when insonifying the whole segment, in particular, at its end point. Therefore, the length estimation of the presented sound segment in the experiment stated in this manner is inefficient. The variance of the trajectory length in subjects' reports can be comparable with the length of the segment at any recording technique, and the number of points within the segment, as well as the fundamental frequency, is of small significance.

Note that the quantization of the virtual trajectory (stepwise changes in amplitudes of the four sources each 4 ms) can cause the vibrato sensation. However, this effect can be observed only with tone signals and the number of points $M < 64$ [3], and when the amplitudes of the tone and modulation are comparable. The stepwise changes in the level of the rectangular signal (at a rate of 250 Hz) that accompany the motion of the sound cursor (305 Hz) are less than 0.5 dB (at $M = 128$). Interference of these signals is insignificant, because the difference frequency is at the edge of both the auditory perception and the frequency band of the used acoustic systems. Therefore, no special measures are needed to provide a continuous shape of the sound signal at $M \geq 128$.

Experiments on recognizing alphanumeric symbols have also shown that it is necessary to highlight the first point of each sound pattern [6]. In this case, the experiment provided 60–90% of correct answers. It is even more significant to highlight the critical points when several sound cursors interact within a single sound pattern (Fig. 6). An increase in the duration of sounding a point from 5 to 16 ms changes the subjective perception of the segment length (100 points) by a factor of two. Increasing the duration leads to a diffuse perception of the edge points; decreasing, to the degeneration of the impression of an extended object to a diffuse pattern.

In the first experiments, we marked the critical points by the sound duration using the repetitive sounding (at a fixed space position) a given number of times. However, this particular solution proved to have a low efficiency for more complex and faster presented sound patterns. The sound duration at critical points that is sufficient for stabilizing the attention was from 40 to 200 ms. For a 100-point segment, the edge points sounded up to 50% of the total time.

Since, in future experiments, we planned to study a possibility to form quasistatic virtual patterns (cursor, scale rules, window borders, and other attributes of the graphic interface), the mode of cyclic presentation of the sound pattern with an opportunity of the reverse presentation was introduced to the sound part of the software interface. The optimum values of certain frequency and time parameters that stabilize the perception of a sound pattern were determined in the cyclic mode using the direct and reverse runs, the result being tested in a single run. For example, at the given frequency and a 4-ms point sounding, when the critical points were not specially highlighted, the cyclic presentation of a straight segment caused a sensation of acoustic beats; however, the effect of horizontal motion was experienced by only a part of subjects (8 persons). When the sound duration at the end points increased to 40 ms, two diffuse regions could be clearly resolved between which the sound motion was perceived. At a 60-ms duration of the edge points, the impression of the motion of the virtual source became more clear, but the variance in perceiving the edge positions increased (it decreased as the sound duration at each point increased to 8 ms).

Therefore, we decided to study a possibility of marking the critical points by simultaneously changing their sounding time and applying a deviation to the fundamental frequency (a stepwise frequency change at the critical point).

Under the same conditions (100 points, 305 Hz), the edge positions of the straight horizontal segment, the minimum sound duration at the critical points was adjusted at 8 ms, while the sound frequency was changed for this time from 305 to 725 Hz in a stepwise manner. This value was much higher than the one nec-
necessary for the perception and efficient localization of the critical point in space at the chosen sound duration and produces a stable and almost identical sensation in all subjects.

It was found that, when the sound duration at the end point of the segment increased (from 8 to 20 ms and longer) and the frequency simultaneously changed, a significant deviation of the segment shape from the linear one was observed. Figure 3 schematically shows a subjectively perceived sound model of a straight segment for different sound durations at the critical points for the frequency deviation $D_2 = 1.3$.

Since a 8-ms duration of the frequency step was sufficient to satisfactorily localize the point (with the detection probability $P > 90\%$), we decided to study the necessary frequency step at the end point versus the frequency of the fundamental tone of the moving horizontal cursor in the absence of the fundamental frequency deviation in both axes: $D_1 = 1.0$ and $D_2 = 1.0$. This problem was solved by measuring the subjective threshold of perceiving a short-time change in the frequency of the moving virtual source of the rectangular signal at the given frequency, velocity, and also duration and localization of the change at the given point of the trajectory in the virtual acoustic plane. In essence, it was necessary to determine a set of parameters of the virtual sound source in the vicinity of the critical points that facilitate the sensory fixation of the respective position and simultaneously do not mask the neighboring points. These requirements were fulfilled with the source parameters increased by a factor of two with respect to their threshold values. These settings reduced the variance in the test results associated with subjects' individual auditory properties. Our results had much greater magnitudes than those obtained in [7] and in audiometric studies of the differential thresholds in resolving the frequency of overlapping tones and impulse sources cited in [10]. The frequency step necessary for the efficient localization of the critical points in space as a function of the fundamental frequency of the horizontally moving sound cursor averaged over tests in 8 subjects is plotted in Fig. 4. Three regions of the fundamental frequency are clearly seen: below 200 Hz, from 200 to 500 Hz, and above 500 Hz. Respectively, in the first region, the effective frequency step is higher than 1.2; in the third one, no higher than 1.02; and, in the intermediate one, it changes within 1.02–1.2. However, these conditions were acceptable when no frequency deviation was applied to the sound cursor in the $Ox$ and $Oy$ directions. For arbitrary sound patterns, these parameters are higher and vary from 1.05 to 1.6. Consequently, it was no less important to determine the relationship between the necessary frequency step at the critical point and the frequency deviation of the sound cursor within the virtual acoustic plane relative to the fundamental tone. The experiments used seven fundamental frequencies ($F_0$: 188, 195, 211, 242, 281, 334, and 585 Hz, and a frequency deviation along the $Ox$ axis from 1.1 to 1.7 for the moving sound cursor.

ACOUSTICAL PHYSICS Vol. 46 No. 1 2000
The test results for one of the subjects are given in Fig. 5. The practically significant $D_t$ is within 1.05–1.6 and depends on the fundamental frequency. The Ox-axis frequency deviation shifts the right end point upwards with respect to the left end point proportional to $D_z$, the frequency step with respect to $F_0 \times D_z$ for the right end point remaining unchanged.

Of course, a more thorough study of the perception process at the critical points of a linear segment, in particular, when the sounding procedure is cyclically reversed, can reveal the differences associated with changing the sign of the frequency deviation and the Ox axis fundamental tone deviation. However, this effect is of no practical importance because the frequency variation at the critical point is substantially greater than the threshold in perceiving the frequencies at two adjacent points. The sign of the frequency deviation at the critical point is much more important, and this sign should be associated with the ordinate of the virtual acoustic plane, because the fundamental frequency deviation is also dependent on this parameter [11, 12]. As a result of the analysis of the experimental data and the refining tests, the critical point parameters for the fundamental tone above 200 Hz (see Figs. 4, 5) were defined by the following conditions:

1. The virtual acoustic plane is divided into two segments at a point with the ordinate $K$.
2. If the critical point ordinate is equal to or less than $K$, the deviation increases the given point frequency by a factor of $H$.
3. If the critical point ordinate is greater than $K$, the deviation decreases the given point frequency by a factor of $L$.

The empirical values of these quantities for the fundamental tone no higher than 550 Hz were found to be $H = 1.2, L = 1.2$, and $K = 43/128$ with the lower left point of the virtual acoustic plane taken as an origin. These rules, which control the parameters of the straight horizontal virtual segment were tested in similar experiments with the segment placed vertically in the acoustic plane. No additional correction or substantial change in the control parameters of the acoustic cursors were necessary.

The frequency deviation (stepwise change) at the critical points of the virtual sound pattern proved to be an efficient control parameter. In the case of a deficiency of time, this parameter is convenient for marking the points necessary for switching the attention of the audio display user.

Figure 6 plots the apparent trajectories of a sound pattern formed by the simultaneous motion of two sound sources when the critical points are (1) not marked, (2) sounded for 40 ms, and (3) sounded for 40 ms and experience a fixed 8-ms-long frequency step. The contraction of the region of uncertain space localization of the perceived virtual sound source trajectory is clearly seen. Unlike the familiar model of interaction between two sound fluxes [13], in which the time factor is the main one for merging and splitting, our model introduces the frequency marking of the critical points to control the subject’s attention.

By dividing the functions switching the attention to a region that requires the space localization into the stepwise frequency change and fixing the attention through a longer sounding of the critical point at the initial frequency, we substantially decreased the sound duration at the critical points and stabilized the perception of the virtual sound pattern borders. The proposed technique provides new means for controlling the space parameters of patterns in virtual acoustic media.
ACKNOWLEDGMENTS

We are grateful to the editorial board and the reviewer for useful remarks, attention, and comprehensive analysis, which helped us in improving the paper.

The work was carried out under the project "Applied Software Package for Transforming a Graphic Image into Structured Acoustic Field for Blind PC Users" (state registration no. 01.960.010307) supported by the State Science and Technology program "Informatization of Russia."

REFERENCES


Translated by A.D. Khzmalyan