EXPERIMENTAL EVALUATION OF AUDITORY DISPLAY AND SONIFICATION OF TEXTURED IMAGES

Antonio Cesar Germano Martins
Laboratório de Sistemas Integráveis (LSI)
Escola Politécnica da Universidade de São Paulo,
Av. Prof. Luciano Gualberto, 158, Trav. 3,
05508-900 - São Paulo - SP - Brasil.
Email : amartins@lsi.usp.br

Rangaraj Mandayam Rangayyan
Department of Electrical and Computer Engineering,
The University of Calgary, Calgary, Alberta,
T2N 1N4, Canada.
Email : ranga@enel.ucalgary.ca

ABSTRACT
In order to verify the potential of proposed auditory display and sonification methods for aural analysis of textured images, a set of experiments was designed and was presented to 10 subjects. The results obtained and limitations of the methods are discussed.

INTRODUCTION
It has been shown that the auditory system is very useful for task monitoring and analysis of multidimensional data [1]. However, the use of sound in scientific data analysis is rather rare, and analysis and presentation of data are done almost exclusively by visual means. Even when data are the result of sounds, such as an ultrasound exam or sonar, they are first mapped to an image and visual analysis is performed.
Sonification of data can be generally represented as in Figure 1. Data characteristics need to be mapped to sound attributes in order to be presented and analyzed. This mapping is an important task, and appropriate design of the mapping function is the difference between a successful sonification procedure or failure.

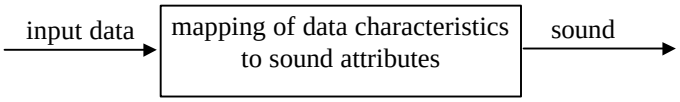

Figure 1 - Sonification process.

Many processes related to our everyday experiences can be used to map data to sound, such as metaphorical and affective association [2]. Common metaphorical associations include high frequency for an increase in a data characteristic or a rapidly changing parameter of the data, and conversely, low frequency for a decrease or slow rate of change. Affective association tries to associate the feelings evoked by a sound with the data controlling the sound. As an example, in a project on the identification of features in magnetic resonance images of the brain, an initial suggestion from our collaborators in the field of radiology was to map pleasant sounds to healthy tissue and unpleasant sounds to diseased areas.
A few mappings from data to sound have been established in the literature [1]. Walker and Kramer [3] argued that one of the considerations that should be taken into account is that a particular mapping choice must rely on the performance of a task, rather than on merely the feeling of the designers that the choice made is intuitive; their preliminary result shows that what is intuitive for some may not be intuitive for others.

We are particularly interested in sonification of textured images [4]. Among the many problems involved in this task, the first that comes to mind is that sound is essentially a change of pressure with time, i.e., sound is something that evolves in time, whereas an image is a static object. How can something static be mapped into something that is time-varying? One way of addressing this problem is to choose a sound signal, for example a sinusoidal oscillator, and associate pixel

characteristics of the image to some of the parameters of the oscillator. Meijer [5] proposed a sonification method using this approach, with the amplitude of a sinusoidal oscillator being proportional to the pixel gray level and the frequency being dependent on the position of the pixel. The image is scanned one column at a time, and the associated sinusoidal oscillator outputs are presented as a sum, followed by a click before the presentation of the next column. The sound signal contains all the information of the image, but analysis can be very difficult. The mapping seems to be "natural", but is arbitrary and needs to be tested with a large group of subjects.

We have developed methods for auditory display and sonification of textured images drawing support from the model for speech generation [6]. Random texture may be modeled as the result of filtering (convolving) a random noise field with a "spot" [7]. Ordered or quasi-periodic texture could be seen as the result of the convolution of a spot (texture element or texton) with an ordered field of impulses. The models compare very well with the models for speech, where we have voiced speech as the result of filtering a quasi-periodic glottal impulse train by the characteristics of the vocal tract, or unvoiced speech due to filtering of a random noise input by the vocal tract. We have drawn an analogy between speech and texture synthesis [4,9] since both can be generally modeled as the result of a convolution of an impulse field with a basic wavelet.
In the case of random texture, we convert the two-dimensional (2D) image into one-dimensional signals by taking projections (Radon transform) of the image at several angles [8,9]. By the Fourier slice theorem [10], we know that the Fourier transform of a projection of an image is equal to the radial slice of the 2D Fourier transform of the image at the angle of the projection. By presenting several projections as sound, we deliver the spectral characteristics of the random texture. The mapping of the projection data to sound has been fully discussed in a previous publication [4]. Linear prediction was used to model the projection data and generate sound signals extended to 0.5 s per projection at a sampling rate of 8 kHz.

For quasi-periodic texture, we map salient attributes of the texture element and periodicity to sound parameters [4]. In particular, we use projections of the texture element as basic wavelets to synthesize voiced-speech-like sounds, with the pitch being a function of the vertical periodicity. The horizontal periodicity is used to provide rhythm in the presentation of the series of projections of the texture element.

The two methods were designed to have close connections to the models for texture synthesis and derived in the context of speech models. We believe that our methods have the desired aspect of natural mappings, but this by itself is not an adequate proof of the concept. In order to verify the potential of the methods for auditory analysis of textured images, we conducted a set of experiments with several subjects. The following sections present details of the experiments and the results obtained.

AUDITORY EXPERIMENTS

The experiments were designed to verify the validity of the model used for the auditory mappings. The main purposes of the experiments were, for the case of random texture, to verify if subjects could:

- classify random texture according to the shape of the spot,
- place in order the sounds derived from images with spots of the same shape but with different spot sizes, and
- associate a textured image with the sound generated via the auditory display procedure proposed;

and in the case of periodic texture to verify if:

- the mappings proposed have a "natural" association with the image,
- the mapping functions give the possibility of ordering images according to variations in their parameters.

We designed a total of 15 experiments, with 10 for random texture and 5 for periodic texture. For random texture, we conducted the following experiments:

- Group 1: Sounds of textures generated with circles of diameter 4, 12, and 20 pixels; squares of side 4, 12, and 20 pixels; ellipses of two major-minor axis ratios; and two hashes of different sizes as the spot were provided. The subjects were asked to place the sounds in two sets. This experiment was designed to evaluate if the sounds could be used to identify the smooth or sharp nature of the spot.

- Group 2: Sounds of textures generated with the same spot (circle or square) but different sizes were provided. The subjects were asked to suggest an order for the sounds. The question here is if the auditory display preserved size or scale information.
- Group 3: Sets of three and four images and their sound signals were provided. The subjects were asked to associate the sounds with the images.
- Group 4: Sounds of textures generated with two spots of different sizes, such as circles of diameter 4 and 12 pixels and the two ellipses were provided. The subjects were asked to place the sounds in two sets.

As the derivative operation emphasizes high-frequency components of a signal, we had suggested in a previous publication [4] that the auditory display of the derivatives of the projections could lead to improved analysis of the sounds. To test this hypothesis, the experiments listed above were repeated with the derivative operation included.

For periodic texture, the subjects were asked to associate an image with a sound when in experiments of:

- Group 5: Three images with the same texton but different horizontal spacing and their sounds were provided.
- Group 6: Three images with the same texton but different vertical spacing and their sounds were provided.
- Group 7: Three images with different texton size and their sounds were provided.
- Group 8: Three images with the same texton size but varying texton shape and their sounds were provided.
- Group 9: Four images with different horizontal and vertical spacing of textons of different size and shape and their sounds were provided.

To present the experiments, we designed World-Wide Web (WWW) pages for use with a web browser. The subjects could conduct the experiments on their own in an easy manner, and listen to the sounds as many times as necessary.

The experiments were conducted by ten subjects of different age, gender, and professional background. The auditory and sonification methods were not explained to the subjects before the experiments. Subjects were asked to not guess, but answer the questions only if they were sure about their judgment. Informed consent was obtained from the subjects, who were free to abandon the experiment whenever they wanted to.

RESULTS AND DISCUSSION

Tables 1-7 show the main results obtained for the case of auditory display of random texture.

Table 1

Group 1	without the derivative	with the derivative
Placed the sounds of the images generated with the two ellipses in the same set	100%	70%
Placed the sounds of the images generated with the two ellipses and the circles with diameter 12 and 20 pixels in the same set	80%	60%
Placed the sounds of the images generated with the two hashes in the same group	90%	80%
Placed the sounds of the images generated with the two hashes and the square of side 12 pixels in the same set	90%	60%
Put the two hashes, the square of side 4 pixels and the circle of diameter 4 pixels in the same group	60%	60%

Table 2

Group 2	without the derivative	with the derivative
Suggested the correct order for the images generated with circles	100%	60%
Suggested the correct order for the images generated with squares	60%	60%

Table 3

Group 3 - spots: circles of diameter 4, 12 and 20 pixels	without the derivative	with the derivative
Suggested the correct association of image to sound for the images generated with the circle of diameter 12 pixels	90%	30%
Suggested the correct association of image to sound for the images generated with the circle of diameter 4 pixels	60%	80%
Suggested the correct association for all the images and sounds	60%	20%

Table 4

Group 3 - spots: ellipse of axes 14 and 20 pixels, ellipse of axes 10 and 20 pixels, and hashes of sides 12 and 20 pixels	without the derivative	with the derivative
Placed the images generated with the hashes and the ellipses in two sets correctly	80%	90%
Suggested the correct association for all the images and sounds	20%	10%

Table 5

Group 3 - spots: ellipse of axes 14 and 20 pixels, ellipse of axes 10 and 20 pixels, and circles of diameter 6 and 10 pixels	without the derivative	with the derivative
Placed the images generated with the circles and the ellipses in two sets correctly	90%	40%
Suggested the correct association for all the images and sounds	40%	10%

Table 6

Group 4 - spots: ellipse of axes 14 and 20 pixels, ellipse of axes 10 and 20 pixels, and circles of diameter 4 and 12 pixels	without the derivative	with the derivative
Placed the images generated with the ellipses in the same set	80%	80%
Suggested the correct separation	40%	10%

Table 7

Group 4 - spots: squares of side 4 and 12 pixels and hashes of side 12 and 20 pixels	without the derivative	with the derivative
Placed the images generated with the hashes in the same set	90%	90%
Placed the images generated with the square of side 12 pixels alone in one set	50%	50%
Suggested the correct separation	40%	30%

Analyzing the results in Tables 1 to 7, we make the following observations:

- From the experiments in Group 1: The sounds generated with the proposed AD method convey the information about the shape of the spot used to create the texture. It can been seen that for very small spots, such as the circle with diameter 4 pixels, the high-frequency components of the input noise image are not sufficiently filtered, which makes difficult shape identification of the spot, leading to misclassification.

- From the experiments in Group 2: In the case of textures obtained with the circles as the spot, the size information of the spots was readily recognized and successfully used for ordering the images. For textures associated with the squares, the larger high-frequency components led to a less efficient analysis.

- From experiments in Group 3: Although the correct image-to-sound association accuracy was between 20% and 60%, it can be seen from Table 3 that 90% of the subjects could make an association between the visual order of the images and the aural order of the sounds. This result can be understood, if it is considered that in visual analysis, investigators are not used to provide a description in terms of frequency. The mismatching of the elements in the visual domain with the elements in the aural domain is again present in the results of Tables 4 and 5, where the sounds associated with similar spots were grouped together but the rate of correct association of image to sound is very low.

- From the experiments in Group 4: The results obtained indicate once more that the auditory display procedure can be used to deliver information about the spot, with the possibility of classifying random texture images. The correct rate of 50% for the segregation of the sound of the texture associated with the square of 12 pixels from those associated with the hashes and the square of side 4 pixels is an indication that the information about the size and shape of the spot may not be independent in the proposed auditory display procedure.

We found that sound presentation without the use of the derivative led to a better performance than with the derivative. This could be due to the fact that the initial sound signals indeed had significant high-frequency energy. The derivative operation resulted in very noisy signals that were difficult to analyze.

With the sonification method for periodic texture we obtained the following performances:

- From the experiment in Group 5: 70% of the subjects suggested the correct association of sounds with images presented with different horizontal spacing of the textons.
- From the experiment in Group 6: 90% of the subjects suggested the correct association of sounds with images presented with different vertical spacing of the textons.
- From the experiment in Group 7: 80% of the subjects could associate the texton size differences with the sounds.
- From the experiment in Group 8: 10% of the subjects suggested the correct association of sounds with texton shape, but 60% identified the sound of the image with the circle as the texton.
- From the experiment in Group 9: 70% of the subjects suggested the correct association of sound with images of different horizontal spacing, vertical spacing, texton size, or texton shape.

The results from the experiments in Group 5 and 7 show that rhythm, pitch, and duration are easily identified with horizontal periodicity, vertical periodicity, and size of the spot, respectively. From the experiment in Group 9 there is an indication that, for the images used, the parameters mentioned above are independent.
The results of the experiments of Group 8 indicate that sounds with significant high frequency components are difficult to analyze.

It should be stressed that all the results presented in this paper were obtained without any preview explanation of the mapping procedure to the subjects since we wanted to verify if the auditory display and sonification methods proposed were "natural". We believe that with specific training, the results can be further improved.

ACKNOWLEDGEMENTS
This work has been supported by grants from CNPq- Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brasil; FAPESP- Fundação de Amparo a Pesquisa do Estado de São Paulo, Brasil; the Natural Sciences and Engineering Research Council (NSERC) of Canada; The University of Calgary Research Grants Committee; FINEP- Financiadora de Estudos e Projetos, grant 56.94.0260.00, Brasil; and PROINTER - Pró Reitoria de Pesquisa da Universidade de São Paulo, Brasil.

References

1. G. Kramer ed. - *Auditory Display: Sonification, Audification, and Auditory Interfaces*. Addison Wesley , Reading, MA, 1994.
2. G. Kramer - *Some organizing principles for representing data with sound*. In Auditory Display: Sonification, Audification, and Auditory Interfaces. Addison Wesley, Readings, MA, 1994, pp. 185-222.
3. N.B. Walker and G. Kramer - *Mappings and metaphors in auditory displays: An experimental assessment*. In Proc. of the Third International Conference on Auditory Display, pp. 71-74, Palo Alto, CA, Nov. 1996.
4. R.M. Rangayyan, A.C.G. Martins, and R.A. Ruschioni - *Aural analysis of image texture via cepstral filtering and sonification***.** In Proc. SPIE: Visual Data Exploration and Analysis III, vol. 2656, pp. 283-294, San Jose, CA, Jan. 1996.
5. P. Meijer - *An experimental system for auditory image representation*. IEEE Transactions on Biomedical Engineering, v.39, no.2, pp.112-121, Feb. 1992.
6. L.R. Rabiner and R.W. Schafer - *Digital Processing of Speech Signals*. Prentice-Hall, Englewood Cliffs, NJ, 1978.
7. J.J. van Wijk - *Spot noise***.** Computer Graphics, v. 25, no. 4, pp. 309-318, Jul. 1991.
8. A.C.G. Martins and R.M. Rangayyan - *Complex cepstral filtering of images and echo removal in the Radon domain*, to appear in Pattern Recognition, 1997.

9. A.C.G. Martins, R.M. Rangayyan, L.A. Portela, E. Amaro Jr., and R.A. Ruschioni - **Auditory display and sonification of textured images**. In Proc. of the Third International Conference on Auditory Display, pp. 9-11, Palo Alto, CA, Nov. 1996.
10. A. Rosenfeld and A.C. Kak - **Digital Picture Processing**. Academic Press, New York, NY, 2nd edition, 1982.