Motion Based Visual to Auditory Substitution

Cameron Elliott, Garabed Tashian, Jeffery Crispo, Eugene Chabot, PhD, and Ying Sun, PhD
Department of Electrical, Computer, and Biomedical Engineering
University of Rhode Island, Kingston, RI 02881-0805 USA

***Abstract*— Various techniques have experimented with sensory substitution as a means to relay information from an impaired sense to another. Often, the limited sensory bandwidth of modalities makes replacement of the visual system difficult or impossible. This study has explored the use of a visual to auditory substitution technique with a reduced set of visual information, namely motion information. Visual information captured through a camera and processed in Matlab is presented in stereo through headphones. Objects moving in the field of view (FOV) have varied audio outputs with high and low pitched tones mapping to larger and smaller motion mass, respectively. The location of the object's center of mass is presented through modulation of stereo audio. Processing techniques have been selected in order to meet real-time, lower power constraints of an assistive technology device.**

I. Introduction

Visual impairment is a problem that affects many with numerous causes, but all can have a significant impact to humans. Over the years, many individuals have attempted to repair the visual system or alternatively supplement the impaired system through substituting another sensory system called sensory substitution. While many techniques have not been able to gain widespread use or acceptance, sensory substitution is an active area of research. Visual sensory substitution is particularly a difficult problem with other sensory systems lacking sufficient informational capacity. While brain plasticity could help supplement informational capacity, such as with reading braille [2], the extent to which it can augment processing is still unknown. Researchers, such as Meijer [1], have developed techniques providing a nearly direct translation of information, but at a significantly reduced resolution and frame rate. Even with this large reduction in detail, this large amount of information almost completely impairs the auditory sense from normal environmental inputs. With a focus on imposing less on the normal auditory sensing abilities, a novel approach has been devised that utilizes a subset of visual information, namely characteristics extracted from motion information, and relay the results through a simple stereo audio algorithm. The current algorithm assumes one object in the field of view, and presents information to infer size, location, and direction of motion of the motion mass using stereo audio amplitude modulation and frequency.

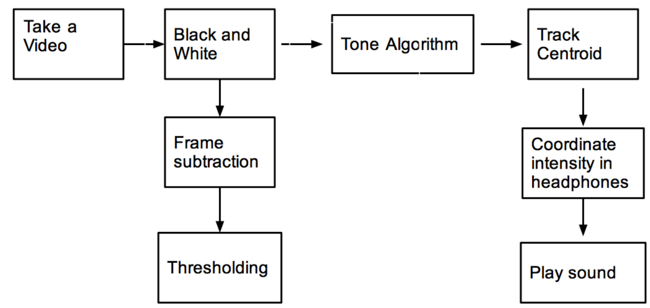

Figure 1: Sequence of sound algorithm

II. Methods

For the proposed device, a number of components are necessary as illustrated in Figure 1. Matlab based scripts have been written to provide the necessary computations. The following sections detail the image processing and audio generation techniques.

A. Image Processing

To start the image processing, color images are captured from a camera. Next, the image is converted to grayscale and subsequent frame pixel intensities are subtracted, similar to the technique detailed by Han [3]. Ideally, the only pixels with a non-zero value would be the portions of the field of view that have changed due to motion. A binary image is generated from all of the pixels exceeding a heuristically set threshold. Further noise reduction is performed through applying a sequence of morphological image operators. A number of noise sources exist, such as sensor and environmental noise, which the effects are reduced by a combination of operators assuming poor spatial organization (i.e. leaves blowing in the wind). From this image, the center of mass is computed to determine the location of the motion, and the number of pixels above the threshold specifies the motion size.

B. Audio Projection

With motion parameters extracted, Matlab is used to create a pseudo-three dimensional sound using stereo audio that the user can interpret to determine where the moving object is located, the direction it is moving in, and the relative size of the object. Depending on the location determined, the sound amplitude for each speaker is set as a fraction of the

Figure 2: Original image of car

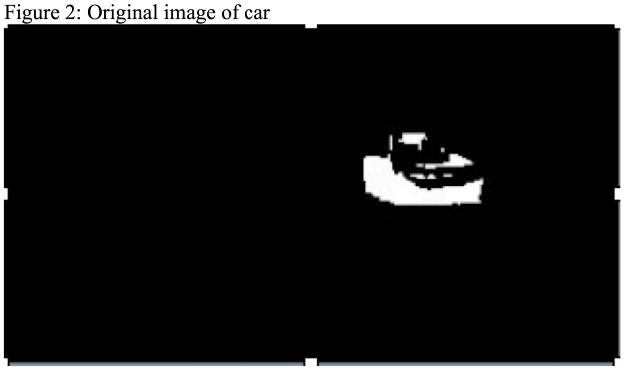

Figure 3: Image after frame subtraction and filtering

distance from the opposite side of the frame. For example, when an object is on the left hand side of the frame, the left speaker would play at full volume, while the right speaker would be off. When the object is in the center of the frame, each speaker would play at half volume. Based upon the change in volume from frame to frame, the user could infer direction of motion.

III. Result

The system has been successful implemented in Matlab on a Macintosh computer. For the present the input is a pre-recorded color or grayscale video clip such as the one shown in Fig. 2. The moving object is detected as shown in Fig. 3. The motion information in the horizontal direction is conveyed to the user by using a sequence of short beeps. The duration of each beep is 0.5 s. The interval between adjacent beeps is a programmable constant, currently set at 2 s. The beeps are turned on when there is a detectable motion. The bearing of the moving object is presented by stereo sound quantized into 21 equally spaced angles. The size of the object is quantized into small, medium, and large. The size is presented by the pitch of the beep: low pitch for small object, medium pitch for medium-size object, and high pitch for large object.

IV. Discussion

The proposed visual to audio substitution method, involving motion extraction and audio presentation algorithms, has shown promising preliminary results in a controlled environment. With the computationally efficient image processing algorithms, as was stated in [4], ego-motion introduces motion artifacts that are difficult to distinguish from object motion, slow moving objects (relative to the frame rate) may not be detectable since frame to frame variations may be minimal, and only one object is assumed to be moving in the frame. The first problem could possibly be solved with a more robust algorithm that could possibly include an image stabilization technique. The second problem of detecting small motion could be addressed by a number of techniques, such as increased camera resolution or slower frame rate. A slower frame rate similarly may be infeasible if the user cannot get the information in a timely manner to make decisions to interact with the changing environment. The last major limitation mentioned relates to number of objects detected in the frame. Our motion detection algorithm assumes that only one object is producing the motion in the field of view. If there are two objects moving in opposite directions, the direction of movement of the largest object would dominate the location presented to the user. While this approach could be enhanced through image segmentation, this increases the computational complexity.

Further work is planned for this technique to test and improve the design. A formal human study is currently in progress to examine the ability of the individuals to extract information related to size, direction, and location of motion. Another effort includes adding more parameters to our motion detection algorithm for the detection of multiple objects. A portable version of the device is also being developed for in-situ testing of performance in real-world environments.

Acknowledgment

This study was supported in part by a grant from the National Institutes of Health (2R44NS048682-02A1, PI: Sun).

References

[1] Meijer, P. "An experimental system for auditory image representations." Biomedical Engineering, IEEE Transactions on, pp. 112-121, 1992.

[2] Sadato, N, Pascual-Leone, A, Grafman, J, Ibanez, V, Deiber, M, Dold, G, and Hallett, M. "Activation of the primary visual cortex by braille reading in blind subjects," Nature, vol. 380, no. 6574, pp. 526-528, 1996.

[3] Han, X., "Active image motion seeking (aims) camera system," Ph.D. dissertation, University of Rhode Island, 2002.

[4] Chabot, E., and Sun, Y.. "Visual-to-Tactile Interface to Detect Motions in Real-time for Persons with Visual Impairments" Proceedings of the 32nd Annual Northeast Bioengineering Conference. pp. 63-64. 20