SONIFICATION METHOD TO ENHANCE THE DIAGNOSIS OF DEMENTIA

Agnieszka Roginska [1], Hariharan Mohanraj [1], James Keary [1], Kent Friedman [2]

[1] Music and Audio Research Lab
Music Technology
New York University
New York, NY
{roginska, hari.m, jpk349}@nyu.edu

[2] Division of Nuclear Medicine Department of Radiology
NYU School of Medicine
New York, NY, USA
kent.friedman@nyumc.org

ABSTRACT

Positron emission tomography (PET) scans of brains result in large datasets that are traditionally analyzed using visual displays and statistical analyses. Due the complexity and multi-dimensionality of the data, there exist many challenges in the interpretation of the scans. This paper describes the use of a sonification method to assist in improving the diagnosis of patients with different levels of Alzheimer's dementia. A triple-tone method is introduced, and the audible beating patterns resulting from the interaction of the three tones is explored as a metric to interpret the data. The sonification method is presented and evaluated using subjective listening tests. Results show the triple-tone sonification method is effective at evaluting PET scan brain data, even for listeners with no medical background.

1. INTRODUCTION

The diagnosis of medical conditions has been transformed with the advancement of medical imagining techniques, including X-rays, magnetic resonance imaging (MRI), and positron emission tomography (PET). These techniques provide physicians with multi-dimensional and time-varying datasets that continue to increase in precision and resolution. Until recently, the data acquired by these imagining techniques has been primarily presented with visual displays, using visual analysis as the principal method of evaluation and diagnosis.

There exist many limitations and challenges in the current representation and analysis of the large amounts of data generated by medical imaging techniques. Two areas that could benefit significantly from improved display and analysis are diagnostic accuracy and inter-observer variability. Although diagnostic accuracy is higher than ever, clinicians still have difficulty detecting certain conditions when the visual analysis of the information provided leads to imperceptible differences between health and disease. A possibility is that visual imagining techniques, no matter how sophisticated, are unable to provide the level of differentiation necessary to detect differences between health and disease. Or perhaps the information is there but is imperceptible to the human eye, even with advanced visual display methods. The second area of needed improvement is inter-observer variability. There exist large diagnostic inconsistencies among clinicians upon visual inspections of medical imaging datasets.

In the recent past, significant efforts and advancements have been made in the exploration of alternative methods and techniques or representation of medical imagining data with the hopes of improving the accuracy and consistency of diagnoses. Quantitative analysis techniques have been used to supplement the visual representation (Piper, 2007) (Nelson, Piper, Friedland, & Freeman, 2007). Advanced computer processing techniques have been explored to find new ways to detect disease, including multi-parametric predictive modeling (Najafi, et al., 2012), and machine learning (Oh, et al., 2012).

In order to further improve the analysis of medical imagining data, researchers are exploring alternative methods of analysis and display, including sonification. This paper describes the use of a sonification method to aid in diagnosing patients with different levels of Alzheimer's dementia (AD). The sonification method is presented and evaluated using subjective listening tests.

2. BACKGROUND

Sonification methods and techniques have been explored for medical applications over the past 20 years, for both aesthetic representations as well as diagnostic tools. Rossiter and Ng (1996) developed a tool to sonify full-body medical scans using a path-based model where different biological materials (soft tissue, bone, fat and air) were assigned to a unique musical instrument.

Sonifications of human electroencephalogram (EEG) have been performed by Hermann et. al. (2002) where three parametric sonification techniques for EEG data were presented. Hermann built upon this work with Baier and Stephani (2007) and developed a sonification technique for EEG data aimed at detecting and categorizing epilepsy, using a multi-channel sound reproduction system in order to exploit the spatial hearing capabilities of the listener.

Ballora et. al. (2004) developed a sonification tool to analyze electrocardiogram (ECG) data of human patients. The ear's sensitivity to rhythmic pattern perception was used to identify and differentiate between various conditions of the heart, and has proven to have diagnostic potential.

Methods for sonifying PET scans of brains with AD were presented by Roginska et al (2013a and 2013b). This research presented a sonification tool for the exploration of augmenting traditional diagnosis methods. The SoniScan++ tool is used for the exploration of sonification methods for the interpretation of brain scans. Users have the ability to define medical image scan paths, sound synthesis methods, mapping between the data and sound attributes, and spatialization method.

The sonification method and study presented in this paper follows on the work presented in Roginska et al (2013a and 2013b). We present a novel sonification method where the brain is segmented into three regions, and each region is mapped to a different frequency. The interaction of the tones of different frequencies results in beating patterns, which are easily perceived by the human ear. The different beating patterns that can be created, can be illustrated mathematically through additive synthesis, where frequencies are added point by point. The basics of additive synthesis tell us that, when two frequencies are added together, an oscillating amplitude envelope is created at a rate that is the difference between those two frequencies; otherwise known as a beat frequency or beat envelope. If the frequencies are very close together, the psychoacoustic phenomenon is understood more on a tonal level, i.e. our brain would interpret these two distinct frequencies as one frequency that is an average of the two, along with the beat envelope around that frequency. If the frequencies are further apart, the psychoacoustic phenomenon is more temporal, as our brains interpret them as distinct frequencies peaking at different times, again with the beat envelope around these frequencies. Adding a third frequency adds two more beat envelopes, and more possibilities for tonal and rhythmic complexities. Thus each brain scan has a unique rhythmic pattern, a signature sound. This research exploits the ear's sensitivity to complex rhythms, as established by Ballora et. al. (2004), through the psychoacoustic phenomenon of beating frequencies.

The sonification method presented here has its basis in the field of medical imaging informatics. In the case of diagnosing different forms of dementia, the metabolic activity in different regions of the brain needs to be measured and compared. AD is characterized by decreased metabolic activity (hypometabolism) in the parietal and temporal lobes of the brain. More severe cases of AD also present hypometabolism in the frontal lobe of the brain (Frackowiak, et al., 1981). Hence, the diagnosis of AD in patients is performed by comparing metabolic activity of the parietal, temporal, and frontal lobes to the metabolic activity of other lobes that are not generally affected by Alzheimer's disease, such as the sensorimotor cortex. PET scans provide a 3-dimensional measurement of metabolic activity in different regions of the brain. This sonification method aims to aurally display the metabolic activity of these lobes of interest.

3. METHODS

3.1. Dataset Preprocessing

The datasets utilized for sonification consisted of 32 de-identified PET/CT scans of human brains diagnosed with varying stages of Alzheimer's disease, obtained from the Radiology Department of New York University Langone Medical Center. These 32 brains scans consisted of 8 brain scans in each of the four categories of diagnosis of AD; Normal, Mild AD, Moderate AD, and Severe AD.

All datasets utilized for sonification were spatially warped to a standard brain model, and hence were all spatially consistent. Spatial normalization is a necessary step in conducting statistical analysis across several brain datasets (Piper 2007). This process was done using the medical imaging software MIM, a data visualization tool for PET scan data (www.mimsoftware.com). In addition to pure visualization tools, MIM Software also contains statistical analysis tools that are capable of segmenting the brain into its various lobes and providing statistical deviations from normalcy based on a standard database of normal brains (Piper 2007).

After spatial normalization all dataset pre-processing and lobe segmentation was performed with SoniScan++, our primary data analysis tool for sonification. SoniScan++ was developed in the C++ programming language for the purpose of this research.

In the case of each brain, a subset of each dataset was chosen for sonification; the 30th lateral slice (from the top) of the spatially normalized dataset. This particular slice of each spatially normalized dataset passes through representative regions of the frontal lobe, parietal lobe, and sensorimotor cortex.

3.2. Lobe segmentation

The spatially normalized sub-datasets were then segmented into the three lobes of interest (frontal and parietal) and the reference lobe (sensorimotor cortex). This segmentation was performed with the aid of the MIM software. MIM performs its own automatic lobe segmentation of the spatially normalized brain datasets and provides the segmentation information to the user in the form of the DICOM standard RTSTRUCT file format. These files contain 3-D contours called Regions of Interest. However, for the sake of time, the RTSTRUCT files were not utilized in their entirety for this work. Only the RTSTRUCT's coordinate points, outlining the general boundaries of the lobes of the spatially normalized sub-dataset, were used. These coordinates were set in SoniScan++ and the contours were approximated to straight lines connecting the coordinates (Figure 1).

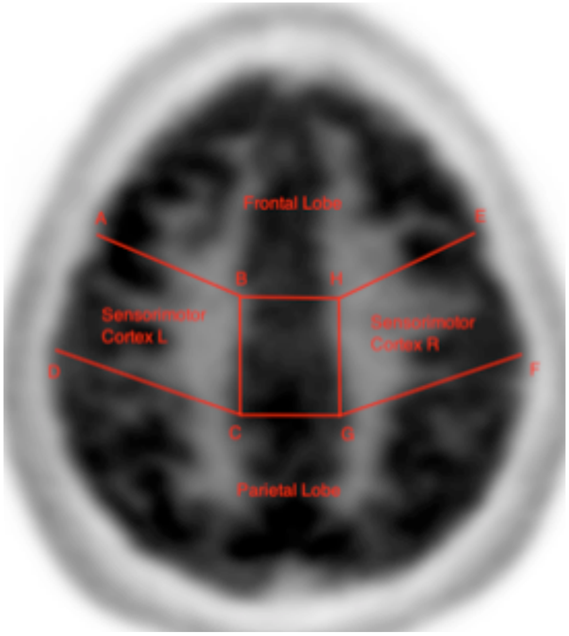

Figure 1 Lobe segmentation of lateral slice

This approach meant that there remained some irrelevant data points within each lobe's segmentation. First, some data points in the lateral slice lie outside the actual brain area. The second category of irrelevant data points consisted of those that lie within the brain, but are medically irrelevant. The brain can for the most part be divided into white matter and gray matter. The relevant metabolic activity for the diagnosis of AD is that of the grey matter of the brain, and hence the white matter content is medically irrelevant and should not be included as data that is sonified. Both issues manifest as voxels with lower intensities compared to the areas of interest. Hence, both issues were tackled with one generalized solution. All voxels whose intensity fell below a certain threshold were masked from being sonified. This threshold was set to 45% of the maximum allowable intensity, considering the bit-depth of the dataset. Hence, for datasets with a bit-depth of 15 bits, the masking threshold is set to 14745.15 out of a maximum allowable value of 32767.

3.3. Triple-tone sonification

In order for a sonification technique to directly target the diagnosis of AD, the difference in metabolic activity between the lobes of interest and the reference lobe was mapped to an easily perceivable auditory parameter. The sonification technique presented here is a "triple-tone sonification", which assigns a triangle wave oscillator to each of the three lobe of the brain to be sonified, specifically each portion of the segmented lobes that fall on the lateral slice.

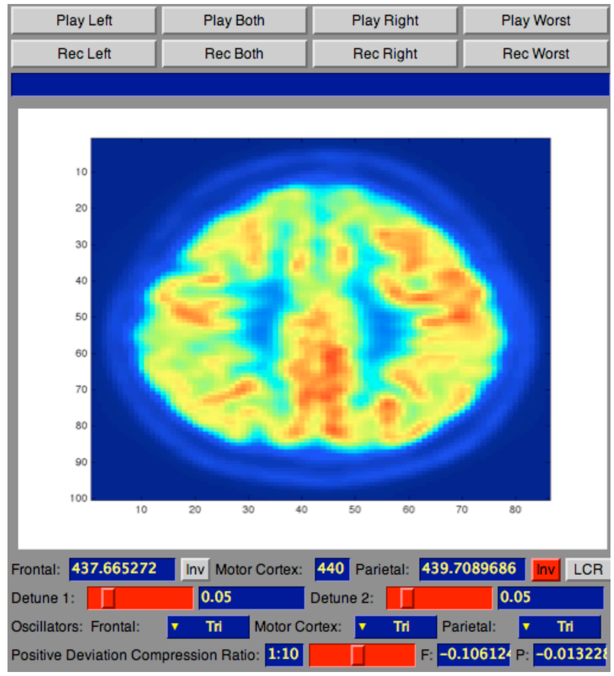

Figure 2 Screenshot of SuperCollider interface for interactive sonification playback and recording

We consider such a technique a high-level approach to analyzing the data, as compared to other techniques performed by this research group in the past (Roginska et. al. 2013). The technique can aim at emulating the mental process of a physician during diagnosis, but presenting the results of this process in the aural domain.

An interactive approach was necessary in order to prototype the technique with versatile functions, to see which auditory parameters best suited the datasets. The sonifications were implemented through the powerful sound synthesis engine and audio programming language, SuperCollider, in the form of a GUI (Figure 2).

3.3.1. Frequency mapping

The frequencies of these oscillators are mapped to the average metabolic activity within each of the three lobes. Hence, differences in metabolic activity between the lobes results in slightly different frequencies of the oscillators. These small differences in oscillation between the three oscillators and triangle wave tones results in beating patterns. The more pronounced the difference between the three frequencies, the faster and more complex the beating pattern becomes. This is directly relevant to the diagnostic method, where higher differences in metabolic activity result in faster beating, more complex rhythms, and eventually splitting of tones.

The frequencies of the tones corresponding to the lobes of the brain under inspection are "detuned" from their default frequency according the deviation of average intensity of the voxels of those lobes from the average intensity of the voxels of the reference lobe(s).

The default frequency used was 440 Hz (A above middle C). The frequency of the frontal lobe was forced to be a positive deviation, and the frequency of the parietal lobe was forced to a negative deviation from the default frequency. Forcing the signs of the deviations of the frontal and parietal lobes ensured that the system would not collapse to two beating tones when there exists abnormalities of equal deviation in both the frontal and parietal lobes. Such a case was undesirable as the complexity of three beating tones was the required artifact to be explored.

3.3.1. Detune factor

The detune factor was used to control the range of frequency deviations given the range of voxel average deviations. The goal of this parameter was to find different levels of beatings to indicate the varying degrees of AD: For example finding the level of beating that would generate split tones indicating severe cases of AD. A higher detune factor would result in a larger frequency deviation, and hence faster beating, for a set of voxel intensity averages.

One approach to determine the detune factor is to make it a constant across cases to ensure complete standardization of frequency deviations. However, preliminary listening indicated that there was not enough differentiability between AD categories when a constant detune factor was used and only highly experienced listeners could easily perceive and categorize the differences in these beating patterns. In order to test the system on non-expert listeners in the future (e.g. physicians), an exaggeration of the audible effect was created. Hence, in order to exaggerate the detuning effect and improve differentiability, an alternative approach to determine the detune factor would be to generate different detune factors based on some feature of the dataset. Although features were identified in preliminary analyses of the datasets, the authors did not have access to a sufficiently large database to make complete generalizations. This research has been reserved for future work. To emulate the desired effect, the detune factor was dynamically assigned to different brains according to the table below.

Brain Type	Detune Factor
Normal brain	0.05
Mild Alzheimer's Disease	0.10
Moderate Alzheimer's Disease	0.15
Severe Alzheimer's Disease	0.20

Table 1 List of detune factor values

Based on these detune factors, the frequencies of the lobes of interest were determined. The relative deviation of average intensity with respect to the average intensity of the sensorimotor cortex (SMC) was linearly mapped to the relative deviation of the oscillator frequency with respect to the oscillator's base frequency through the detune factor coefficient. This can be represented as follows.

$$f_{FL} = f_{default} * \left(1 + DF * \left|\frac{av_{FL} - av_{SMC}}{av_{SMC}}\right|\right) = f_{default} * (1 + DF * |\Delta_{FL}|)$$

$$f_{PL} = f_{default} * \left(1 - DF * \left|\frac{av_{PL} - av_{SMC}}{av_{SMC}}\right|\right) = f_{default} * (1 - DF * |\Delta_{PL}|)$$

where f = frequency, av = voxel average, DF = detune factor, FL = frontal lobe, PL = parietal lobe, SMC = sensorimotor cortex.

3.3.2. Positive compression

We found that, in most cases of diseased brain datasets, the frontal lobe and parietal lobe averages fell below the average of the SMC. However, in several cases of normal brain datasets, the frontal and parietal lobes possessed averages higher than that of the SMC. This does not represent any abnormality with respect to AD, but manifests itself as an abnormality in the sonifications, as now the frequencies of these lobes would be correspondingly higher or lower than that of the SMC.

In order to differentiate between true abnormality (arising from a lower average value of frontal or parietal lobe activity) and misrepresented abnormality (arising from a higher average of frontal or parietal lobe activity), a "compressor" was applied to all average voxel intensity values before performing frequency mapping. All positive deviations from the average of the SMC were compressed by a 10:1 ratio in all cases presented in this work.

4. EVALUTATION

4.1. Procedure

The goals of the evaluation were threefold. First, to evaluate the effectiveness of the triple-tone sonification technique in accurate distinguishability between brains of different levels of Alzheimer's disease. Second, to evaluate the intra-reader consistency of diagnosis. Third, to investigate whether a finer gradation of categorization would result in more accurate or consistent results.

In traditional methods of analysis, the diagnosis is typically divided into a categorization of one of four different levels – normal, mild, moderate, and severe. In addition to using the four- step categorization method, we evaluate a finer, seven-step, scale of categorization. The evaluation was divided into two sections: coarse categorization and fine categorization sections. The task of the participant in each section was to categorize the presented sonification into one of four and one of seven categories in the coarse and fine categorizations respectively.

Prior to the test, the participants were presented with training sonifications for the purpose of familiarizing themselves with the nature of each category. Categories 1, 2, 3, and 4 were chosen to correspond to the diagnoses of normal, mild AD, moderate AD, and severe AD respectively. Ten sonifications were generated for each category, resulting in a total of 40 training sonifications.

The training sonifications were generated through the statistical analysis of the 32 brain scans that were used for trials. The frequencies for each training sonification were sampled from a uniform distribution centered on the mean frequency of the corresponding lobe in the corresponding category, with a range of half a standard deviation in either direction. The sonifications used for the training were excluded from the data set presented during the listening experiment.

In the coarse categorization section, the participant was asked to categorize each sonification into one of the four categories (1, 2, 3, or 4), based on the training sonifications. In the fine categorization section, the participant was asked to categorize each sonification into one of the following categories: 1, 1.5, 2, 2.5, 3, 3.5, or 4. Here, the participant was instructed to rate cases that aligned with the training cases to the integer-valued categories, and the cases that may be interpreted as lying between training case categories as the fractional-value categories.

4.2. Datasets

The evaluation used the 32 unique de-identified datasets of patients with varying severity of Alzheimer's disease, obtained from the NYU Langone Medical Center. The collection consisted of 8 datasets in each of the four categories of Alzheimer's disease severity – normal, mild, moderate, and severe. The "ground truth" diagnosis was performed by Dr. Kent P. Friedman utilizing visualization and statistical analysis tools provided by MIM software.

In each session, two instances of each dataset's sonification were presented to allow for the testing of intra-reader consistency, resulting in a total of 64 sonifications per sessions. The order of presentation of the sonifications was randomized consistently across participants. The datasets were consistently randomized in each listening session to minimize recall bias.

4.3. Participants

The subjective listening test was presented to five participants, all graduate students and faculty members of the Music Technology program, Steinhardt, New York University. This control group was chosen to first evaluate the triple-tone sonification technique because of their musically trained ears, providing a validation before proceeding to evaluation with physicians. All medical aspects of this sonification technique were withheld from the participants, and their evaluation of this technique was influenced solely using auditory parameters.

5. RESULTS

5.1. Response accuracy

The accuracy of response against ground truth was computed for each ground truth category for each participant. The accuracy is computed according to:

$$accuracy = \frac{num\ correct\ responses}{total\ num\ responses}$$

In the case of coarse categorization, a correct response is one where the participant's categorization exactly matches the ground truth of the test case. In the case of fine categorization, a correct response is one that either matches the ground truth, or lies at a distance 0.5 from the ground truth categorization. The results are presented as percentage accuracy for each participant, as well as the mean accuracy for all participants in Table 2 for the coarse categorization, and in Table 3 for the fine categorization.

Participant	**Cat. 1**	**Cat. 2**	**Cat. 3**	**Cat. 4**	**Mean**
P1	100%	88%	81%	88%	**89%**
P2	88%	94%	38%	94%	**78%**
P3	94%	100%	81%	100%	**94%**
P4	100%	94%	75%	100%	**92%**
P5	94%	94%	44%	100%	**83%**
Mean	**95%**	**94%**	**64%**	**96%**	**87%**

Table 2 Accuracy of participants in coarse categorization for categories 1, 2, 3 and 4.

Participant	Cat. 1	Cat. 2	Cat. 3	Cat. 4	**Mean**
P1	88%	88%	81%	88%	**86%**
P2	100%	100%	56%	100%	**89%**
P3	100%	100%	100%	100%	**100%**
P4	94%	100%	88%	100%	**95%**
P5	100%	100%	69%	100%	**92%**
Mean	**96%**	**98%**	**79%**	**98%**	**93%**

Table 3 Accuracy of participants in fine categorization

5.2. Response consistency

Each test case as presented to the participant twice in each section to allow for the testing of intra-reader consistency. In the case of coarse categorization, the responses to a pair of duplicate test cases are said to be consistent if both cases were given the same response by the participant. In the case of fine consistency, the responses to a pair of test cases are said to be consistent if both cases were given responses that differ by no more than 0.5. The percentage of the 32 case pairs that were consistent is the consistency percentage. The consistency percentages for each participant in both coarse and fine categorization are presented in Table 4 below.

Participant	Coarse Categorization	Fine Categorization
P1	84%	88%
P2	84%	100%
P3	94%	100%
P4	97%	88%
P5	84%	88%
Mean	**89%**	**93%**

Table 4 Consistency percentage for coarse and fine categorization

5.3. Fine vs. coarse categorization

A side-by-side comparison of participant accuracies is given below for coarse and fine categorization sections.

Participant	Coarse Categorization	Fine Categorization
P1	89%	86%
P2	78%	89%
P3	94%	100%
P4	92%	95%
P5	83%	92%
Mean	**87%**	**93%**

Table 5 Side-by-side comparison of participant accuracy for coarse vs fine categorization

In the case of coarse categorization, mimicking the diagnosis procedure of AD, participants displayed an overall categorization accuracy of 87%. Four out of five participants (all except participant P1) show an increase in categorization accuracy when allowed a finer gradation. Overall accuracy also increased from 87.19% to 92.5% when allowed a finer gradation.

6. DISCUSSION & CONCLUSIONS

This paper presents a novel triple-tone sonification technique to analyze PET scans of brains with different levels of Alzheimer's dementia. The method was presented and evaluated using subjective testing. Participants involved in the study had no medical experience, but were professional musicians with a trained musical ear.

Results of the evaluation of this sonification method indicate that participants with musically trained ears, are able to categorize the presented sonifications using the triple-tone technique with an average accuracy of 87% using coarse categorization, and an accuracy of 93% when a finer categorization scale was used. This test was comparing the results of the subjective listeners to a baseline diagnosis of a highly experienced medical physician.

The overall accuracy of diagnostic categorization improved with a finer gradation of categorization. This indicates that there exist sonifications generated by this technique that place a brain scan "in between" two categories when evaluated by a listener. In the case of coarse categorization, the in-between scans were perceptually quantized into one of the coarse categories by the participant. In the case of fine categorization, the participant was able to successfully categorize the sonification as an in-between case.

The next steps of this research has been to involve physicians in assessing the benefit of adding the sonification method presented here to the traditional methods of diagnosis, including visual and statistical analysis. This research is currently ongoing. Preliminary analysis of the test data is showing promising results.

7. ACKNOWLEDGMENT

This work was supported by the NYU CTSI Award Number NIH/NCATS UL1 TR000038.

8. REFERENCES

Baier, G., Hermann, T. (2004) The Sonification of Rhythms in Human Electroencephalogram. Proceedings of ICAD 2004, July 6-9, Sydney (2004).

Baier, G., Hermann, T., Stephani, U. (2007) Multi-channel sonification of human EEG, Proceedings of ICAD 2007, June 26-29, Montreal (2007)

Ballora, M., Pennycook, B., Ivanov, P. C., Glass, L., & Goldberger, A. L. (2004). Heart Rate Sonification: A New Approach to Medical Diagnosis. Leonardo , 37 (1) . MIT Press.

Frackowiak, R. S., Pozzilli, C., Legg, N. J., Du Boulay, G. H., Marshall, J., Lenzi, G. L., et al. (1981). Regional Cerebral Oxygen Supply and Utilization in Dementia. Brain (104), 753-778.

Hermann, T., Nattkemper, T. W., & Ritter, H. S. (2000). Sonification of Multi-Channel Image Data. Proc. of the Mathematical and Engineering Techniques in Medical and Biological Sciences (pp. 745-750). CSREA Press

Najafi, M., Soltanian-Zadeh, H., Jafari-Khouzani, K., Scarpace, L., & Mikkelsen, T. (2012). Prediction of Glioblastoma Multiform Response to Bevacizumab Treatment Using Multi-Parametric MRI. PLoS ONE , 7 (1).

Nelson, A. S., Piper, J. W., Friedland, R. P., & Freeman, B. (2007). Probabilistic human brain atlas for functional imaging: Comparison to single brain atlases. The Journal of Nuclear Medicine , 48.

Oh, J., Wang, Y., Apte, A., & Deasy, J. (2012). A Statistical and Machine Learning-Based Tool for Modeling and Visualization of Radiotherapy Treatment Outcomes. Medical Physics , 39 (6), 3763.

Piper, J. W. (2007). Validation of an Artificial Neural Network Computer-Aided Diagnosis Scheme for Alzheimer's Disease using FDG-PET. Wake Forest University, Department of Computer Science.

Roginska, A., Friedman, K. P., & Mohanraj, H. (2013a). Exploring Sonification for Augmenting Brain Scan Data. International Conference on Auditory Display. Poland.

Roginska, A., Mohanraj, H., Ballora, M., & Friedman, K. P. (2013b). Immersive Sonification for Displaying Brain Scan Data. Health Informatics Conference. Barcelona.