5-2013

Quantitative Techniques for PET/CT: A Clinical Assessment of the Impact of PSF and TOF

Joshua Donald Schaefferkoetter
jschaeff@utk.edu

Recommended Citation

Schaefferkoetter, Joshua Donald, "Quantitative Techniques for PET/CT: A Clinical Assessment of the Impact of PSF and TOF. " PhD diss., University of Tennessee, 2013.
https://trace.tennessee.edu/utk_graddiss/1777

To the Graduate Council:

I am submitting herewith a dissertation written by Joshua Donald Schaefferkoetter entitled "Quantitative Techniques for PET/CT: A Clinical Assessment of the Impact of PSF and TOF." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Physics.

Mike W. Guidry, Major Professor

We have read this dissertation and recommend its acceptance:

Hanno Weitering, David Townsend, Jonathan Wall

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Quantitative Techniques for PET/CT: A Clinical Assessment of the Impact of PSF and TOF

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Joshua Donald Schaefferkoetter
May 2013

Acknowledgements

I would first like to thank the primary members of my PhD committee, Mike Guidry, Hanno Weitering, Jonathan Wall, and David Townsend for their continued support throughout my graduate studies. The fifth member, Georges El Fakhri, was essential in defining the course of the research presented in this dissertation and deserves much credit for this work. Bjoern Jakoby was a graduate Physics student working with the molecular imaging group when I joined in 2008; he was largely responsible for helping me get started on this academic path. I would also like to take this opportunity to acknowledge the help of several of the many brilliant scientists working at Siemens Healthcare, who always made time to answer my questions, and without which, this work would not have been possible. These include: Harold Rothfuss, for teaching me the majority of my knowledge about the structure and organization of PET data, Jim Hamill and Judson Jones for their time and assistance with various pieces of software code, Maurizio Conti for numerous helpful discussions about TOF-PET, and Mike Casey, the director of the Physics group, for his direction and many discussions on a wide range of topics. Finally, I would like to thank the three physicians who assisted with the observer study for their valuable time, Drs. Yitong Fu, Karen Wells, and Yong Bradley.

Abstract

Tomographic reconstruction has been a challenge for many imaging applications, and it is particularly problematic for count-limited modalities such as Positron Emission Tomography (PET). Recent advances in PET, including the incorporation of time-of-flight (TOF) information and modeling the variation of the point response across the imaging field (PSF), have resulted in significant improvements in image quality. While the effects of these techniques have been characterized with simulations and mathematical modeling, there has been relatively little work investigating the potential impact of such methods in the clinical setting. The objective of this work is to quantify these techniques in the context of realistic lesion detection and localization tasks for a medical environment. Mathematical observers are used to first identify optimal reconstruction parameters and then later to evaluate the performance of the reconstructions. The effect on the reconstruction algorithms is then evaluated for various patient sizes and imaging conditions. The findings for the mathematical observers are compared to, and validated by, the performance of three experienced nuclear medicine physicians completing the same task.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Definition
AFROC	Alternative free-response receiver operating characteristic
AUC	Area under the ROC curve
BGO	Bismuth germanate
BMI	Body mass index
BKE	Background known exactly
CHO	Channelized Hotelling observer
CT	Computed tomography
DDOG	Dense difference-of-Gaussian
DIFT	Direct inverse Fourier transform
DOG	Difference-of-Gaussian
DWS	Delayed-window subtraction
EM	Expectation maximization
^{18}F-FDG	2-deoxy-2-(^{18}F)fluoro-D-glucose
FBP	Filtered backprojection
FORE	Fourier rebinning
FOV	Field of view
FROC	Free-response receiver operating characteristic
FWHM	Full width at half maximum
LLD	Lower-level energy discriminator
LOR	Line of response
LROC	Localization receiver operating characteristic
LSO	Cerium-doped lutetium oxyorthosilicate
MGH	Massachusetts General Hospital
MLEM	Maximum-likelihood expectation-maximization
MRI	Magnetic resonance imaging
NEC	Noise-equivalent count rate
NPWMF	Non-prewhitening matched filter
OP	Ordinary Poisson
OSEM	Ordered subset expectation-maximization
PET	Positron emission tomography
PMT	Photomultiplier tube
PSF	Point spread function
PWMF	Prewhitening matched filter
ROI	Region of interest
ROC	Receiver operating characteristic
SDOG	Sparse difference-of-Gaussian

SKE	Signal-known-exactly
SKS	Signal-known-statistically
SNR	Signal-to-noise ratio
SSS	Single scatter simulation
SUV	Standard uptake value
TOF	Time-of-flight
WB	Whole body

Chapter 1

PET Background

Introduction

Medical imaging exists for the noninvasive, *in vivo* study of disease in the body. Modalities such as computed tomography (CT) and magnetic resonance imaging (MRI) do this by providing information about patient anatomy, so physicians can identify physical abnormalities in the tissue morphology. In contrast, molecular imaging modalities like positron emission tomography (PET) do not describe the anatomy directly, but rather track specific biological processes. This enables the investigation of bodily function, allowing the identification of abnormalities in tissue physiology.

Central to molecular imaging is the molecule, or more specifically, the radiopharmaceutical (radio-labeled tracer). The tracer is injected into the subject, where it then distributes itself, following a certain pathway. PET is used in the clinic to study many processes including blood flow, tissue perfusion, neurological function, cellular proliferation, and tumor metabolism, and each application uses a unique radiotracer that has been engineered to track a specific molecule or biochemical process.

History

The idea for using positron imaging for clinical use first came in the early 1950's from the necessity to improve nuclear image quality for the detection of tumors. William Sweet, the Chief of the Neurosurgical Service at the Massachusetts General Hospital, was unable to help a female patient with a neurological problem that left her unable to read. He consulted Gordon Brownell, the head of the Physics Research Laboratory at MGH.

Brownell suggested that the use of annihilation radiation from positron emission could increase sensitivity and resolution in brain images (Brownell and Sweet 1953). A simple rectilinear positron scanner, consisting of two sodium iodide detectors, was built, designed for the dedicated task of imaging brain activity *in vivo*.

In the early 1960's, a group at Brookhaven National Laboratory constructed the first true transaxial positron tomograph consisting of a detector ring much like the detection configurations of modern-day PET scanners. This system produced poor images because of insufficient data sampling and inadequate reconstruction methods (Schlyer 2004).

After the invention of X-ray computed tomography in the 1970's (Hounsfield 1973), more sophisticated reconstruction techniques were developed. This opened the door to more advanced reconstruction techniques that would ultimately be applied to positron imaging. These new methods were implemented by Michael Phelps in the mid 70's at Washington University in St. Louis (Phelps, Hoffman et al. 1975). Phelps was the first to show the use of PET as a non-invasive tool for quantifying regional glucose metabolic rates in the brain (Phelps, Huang et al. 1979) and how different parts of the brain are active during various mental and physical tasks. (Phelps, E. et al. 1981) Phelps and Hoffman were responsible for the ECAT tomograph, and have since been involved in the development of four generations of PET scanners. Phelps is widely recognized as the inventor of PET.

Radiotracers

The utility of PET depends on the tracer being imaged; each has its own specific application. The most widely used radioisotopes in PET are fluorine-18 (^{18}F), carbon-11 (^{11}C), nitrogen-13 (^{13}N) and oxygen-15 (^{15}O) because of the relative ease of their substitution. Table 1 lists the properties of some commonly used isotopes.

Table 1. Examples of commonly used isotopes used for PET

Isotope	***Half life***	***B^+ Energy (keV)***	
		Mean	***Max***
^{11}C	20.38 min	390	970
^{13}N	9.97 min	490	1190
^{15}O	2.03 min	740	1720
^{18}F	109.77 min	250	635

These isotopes are formed in a cyclotron, where a high energy proton beam bombards a specific target, creating the unstable, neutron deficient isotopes. In the cases of ^{11}C, ^{13}N, and ^{15}O, the newly formed isotopes can directly replace the stable atoms ^{12}C, ^{14}N, and ^{16}O that commonly exist in organic molecules. However, because fluorine is not naturally occurring in biomolecules, ^{18}F often replaces a hydrogen atom or a hydroxyl group of an otherwise naturally occurring compound. In the case of deoxyglucose, an analog of glucose, the fluorine isotope substitutes for the 2' hydroxyl group to form ^{18}F-FDG.

^{18}F-FDG, or more specifically, 2-deoxy-2-(^{18}F)fluoro-D-glucose (referred to as FDG for the remainder of this work), is the most common radiopharmaceutical used in PET today. First developed in the 1970s at Brookhaven National Lab, FDG is an analog of glucose in which one of the hydroxyl group at the 2' position in the glucose molecule has been replaced with ^{18}F. FDG is chemically similar to glucose, and so it begins to follow the same metabolic pathway. It is carried into cells by the membrane transport protein GLUT1. Once inside the cell it is phosphorylated by the enzyme hexokinase into ^{18}F-FDG-6-phosphate. Once phosphorylated, the molecule cannot leave the cell, and the absence of the 2' hydroxyl group prevents it from undergoing further glycolysis. It becomes trapped within the cell until the ^{18}F isotope decays into $^{18}O^{-}$, where it then finds a proton (H^{+}) in the aqueous environment to form the missing, but now "heavy", hydroxyl group, and glycolysis continues as normal, resulting in non-radioactive end products. The accumulation of FDG is subsequently used as an index of tissue glucose metabolism.

An FDG-PET scan performed in the appropriate time window (usually 1-2 hours post-injection) gives a good reflection of cellular glucose uptake and phosphorylation within a subject. The increased energy demands of rapidly growing cancers (Warburg, Posener et al. 1931) result in elevated glucose uptake in lesions, thereby making them visible in an FDG-PET image.

The time course of biomarker retention in tissue is quantified differently for various tracers; in the case of FDG, it is modeled as a 2-compartmental system, a central component which remains in equilibrium with plasma concentration, the non-phosphorylated FDG, and a peripheral compartment characterized by irreversible uptake after phosphorylation. In this case, the kinetic behavior of the tracer can be modeled by the following relationship (Patlak, Blasberg et al. 1983).

Equation (1.1) describes the accumulation of tracer in tissue as a function of time.

$$A(t) = K \int C_p(t)\, dt + V_0 C_p(t) \tag{1.1}$$

Here, *A* is the total amount of tracer in the region of interest, C_p is the tracer concentration in the plasma, V_0 is the distribution volume of the tracer in the central compartment, and *K* is the flow rate constant of FDG plasma clearance. The central component is the second term on the right hand side and the peripheral is the first. The quantities *A*, C_p and V_0 are measureable, and solving for the rate of FDG phosphorylation *K* gives important information about tissue metabolism. Dividing both sides by $C_p(t)$ yields

$$\frac{A(t)}{C_p(t)} = K \frac{\int C_p(t)\, dt}{C_p(t)} + V_0 \tag{1.2}$$

Plotting A/C_p vs. $\int C_p dt/C_p$ we arrive at a straight line with slope *K* and intercept V_0. This model is convenient because it uses linear regression to simplify the pharmacokinetic exchange rate analysis.

PET Scanners

A modern PET scanner consists of a horizontal bed which passes through a circular bore encasing multiple rings of detectors. The port diameter of a typical gantry is 70-80 cm with an axial field of view (FOV) around 20 cm. The tomograph is capable of imaging the whole body with a “step-and-shoot” approach, scanning sequential bed positions and then merging the reconstructed volumes. The subject is positioned appropriately, and each bed position is scanned for 1-3 minutes.[1]

The basic function of the PET tomograph is to extract location information from positron radiation. Today's PET detection systems are scintillation-based; once a 511keV photon strikes a detector crystal, it is converted into light. This light is collected by a system of coupled photomultiplier tubes (PMTs), which then output an electrical pulse. The strength of this pulse determines the initial energy deposited in the crystal and can be used to reject (scatter) events that do not satisfy the energy threshold. Scanner performance depends strongly on the properties of the scintillating material; Table 2 lists the properties of various scintillators. The Biograph TruePoint TrueV PET/CT scanner (Siemens Molecular Imaging) employs four rings of 48 detector blocks, each comprising an array of 13 x 13 LSO crystals (Jakoby, Bercier et al. 2008).

Table 2. Comparison of different PET scintillators and their properties

Scintillator	***Light output (photons MeV^{-1})***	***Density (g cm^{-3})***	***Attenuation length (mm)***	***Decay constant (ns)***	***Energy resolution (%)***
NaI	41000	3.7	29	230	7
$Bi_4Ge_3O_{12}$ (BGO)	9000	7.1	11	300	21
Lu_2SiO_5:Ce (LSO)	26000	7.4	12	40	12
Cd_2SiO_5:Ce (GSO)	8000	6.7	14	60	14
BaF_2	1400	4.9	23	0.8	10

---

[1] A routine whole-body scan (from the cranial orbits to mid-thigh) typically includes 4-6 bed positions. Melanoma scans of the entire body use as many as twelve.

An early PET scanner incorporated a single array of 48 NaI crystals, arranged in a hexagonal configuration, yielding a transaxial spatial resolution of 7-15mm (Hoffman, Phelps et al. 1976). Today's PET scanners have over 30,000 LSO or LYSO detector elements, arranged in multiple rings, with spatial resolution around 4mm.

Bismuth Germanate, or BGO, was introduced into PET in the late 1970s; although it does not output as much light as its NaI predecessor, it has a higher stopping power, resulting in increased sensitivity. The need for better spatial resolution called for smaller scintillating crystal elements. The complexity of packing large numbers of small crystals individually bonded to 1" PMTs was overcome in a clever way with the invention of the block detector using PMT light sharing. Introduced in 1986, the first block detector multiplexed 32 crystal elements to four PMTs (Casey and Nutt 1986). The light output shared among the four-PMT array for each crystal element was unique and acted as an identification signature. In this way, the detector packing fraction (closeness with which the detector elements can be packed) could be increased, while the number of PMTs could be reduced.

In the early years of PET when scanners first started incorporating multiple detector rings, 2D acquisition was used; annuli of lead shielding, or septa, were positioned between each detector ring to "shadow" the activity arising outside of the direct plane, thus decreasing the number of random and scattered events and simplifying the reconstruction process. The early 1990s saw the PET scanner move from 2D to optional 3D acquisition, by using retractable inter-ring septa. Scanning in 3D greatly increases the system sensitivity and has the potential to provide higher quality images with shorter patient scan time. Increased sensitivity to true coincident events, however, comes at the cost of increased sensitivity to scattered and random coincidences. Thus, 3D scatter corrections and new reconstruction algorithms were needed.

One of the most important developments of the last decade was the combination of PET and X-ray CT into a single clinical scanner (Townsend, Beyer et al. 1998). This new combined modality provided the ability to use the CT transmission scan for attenuation correction of the PET data, eliminating the need for a lengthy PET transmission scan. The CT transmission data are also used for other corrections prior to, and within, the reconstruction process. Another important benefit of the combination is the ability to obtain information about tissue anatomy and physiology by acquiring superimposed, or “fused”, images.

The first fused images came from individual systems operated independently from each other and, if required, the images could be registered manually using software. As shown in Figure 1, the first prototype PET/CT scanner comprised PET and CT components mounted on the same rotating support within the gantry. The CT and PET imaging fields were separated by 80 cm axially (Townsend, Beyer et al. 2003). Software now automates both acquisition processes, as well as data correction, reconstruction, and co-registration of the final images.

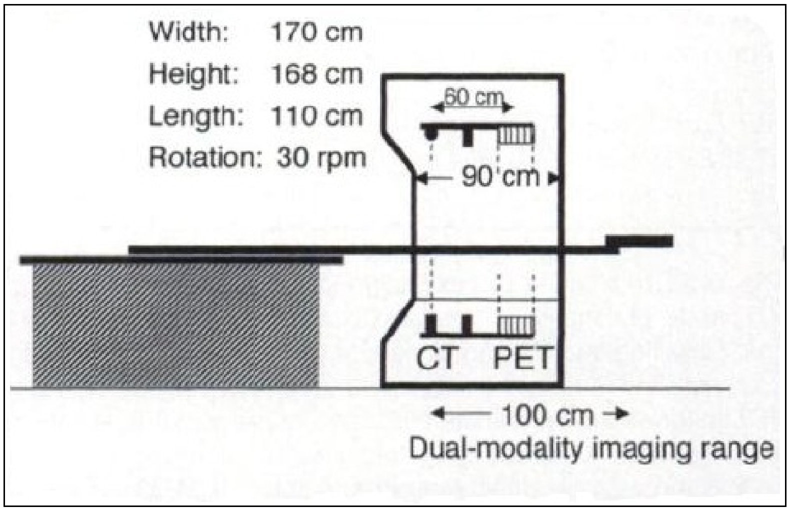


Figure 1. Schematic diagram of early PET/CT concept (Beyer, Townsend et al. 2000)

PET as a Clinical Tool

Many developments have facilitated the integration of PET into the clinic, but few as rapidly as the addition of the CT modality, and now the hybrid PET/CT. In modern PET/CT, the CT transmission scan is integral to important calculations in the processing of the PET data. Additionally, this configuration provides information about anatomy as well as physiology. The diagnostic capability of the combined modality provides an unsurpassed level of patient care. Physicians are able to better diagnose disease, plan treatment, and monitor response more effectively.

PET/CT has revolutionized healthcare, and it continues to expand its utility to a wide array of applications. This is illustrated below in Figure 2 by the growing number of world-wide scanner sales. In 2002, only a couple of years after its introduction, PET/CT systems accounted for nearly half of the total PET scanner sales, with the other half being that of dedicated PET scanners. By 2005, PET/CT systems had almost completely taken over the market (Townsend 2008).

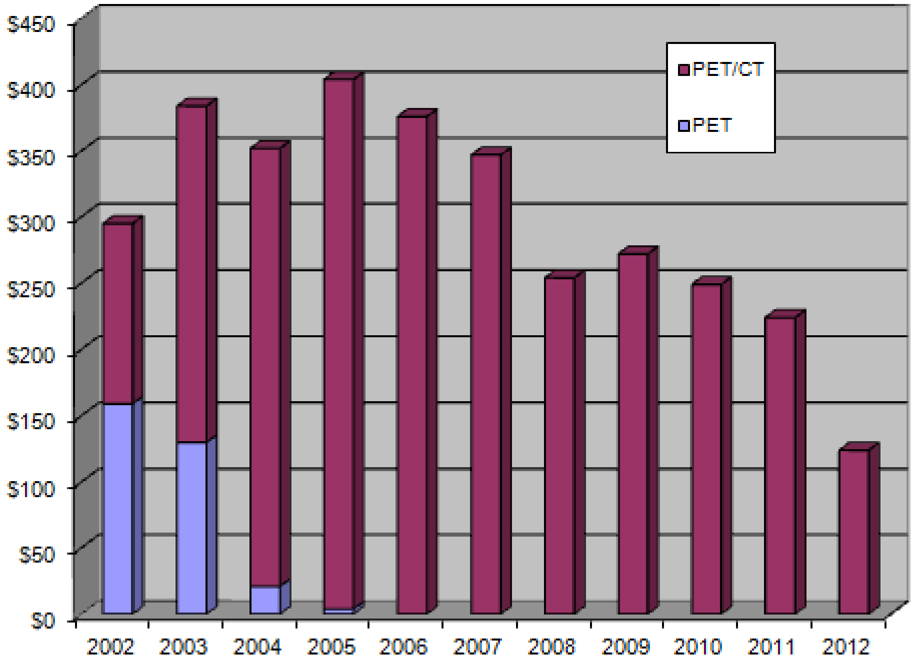


Figure 2. Shipments of PET and PET/CT scanners for the US markets as recorded by the Nuclear Equipment Manufacturers Association

The first PET scans were conducted in the 70’s at Washington University in St. Louis by Michael Phelps, et al (Phelps, Hoffman et al. 1975; Phelps, Hoffman et al. 1976). With the development of FDG later that decade, PET gained more attention for its potential clinical value; but due to the difficulties of the imaging process and the high costs of the radiopharmaceuticals, PET was strictly used for research. However, as more PET studies were performed, its clinical utility began to emerge, specifically in the evaluation of disorders of the heart and brain. This led to the first reimbursement of PET in 1995, for myocardial perfusion imaging using Rubidium-82 (Workman and Coleman 2006). The first Medicare reimbursement of FDG-PET scans in oncology came in January of 1998 for the initial staging of non small-cell lung cancer and the characterization of single pulmonary nodules (Bietendorf 2004). Coverage for colorectal cancer, lymphoma, melanoma, and many others followed in the next few years. The figures below illustrate the improved diagnostic capacity of fused images.

Figure 3. Transaxial PET, CT, and fused images taken from a 68 year-old male diagnosed with cholangiocarcinoma. The combined modality provides physiological information as well as anatomical location. The disease is not obvious in the CT image alone but the PET scan shows an intense focus of metabolic activity medial to the hepatic caudate lobe likely compatible with recurrent malignancy.

Figure 4. Transaxial PET, CT, and fused images of a 40 year-old female with a history of heavy smoking. The intense metabolic uptake illustrated by the PET scan is accompanied by a 2.2 x 2.8 cm upper lobe mass on the CT image, which was later diagnosed as non-small cell carcinoma.

Today, Medicare provides reimbursement of FDG-PET for staging and, more recently, for monitoring of nearly every form of cancer (except prostate). Although a fairly recent development, PET/CT in oncology is now the primary imaging modality, and the continued improvement of scanner performance promises further clinical integration. However, the full potential of PET remains unrealized largely due to the relative limited availability of useful radiotracers. As new tracers are developed and scanning protocols are further refined, the trend in patient care will become more personalized and tailored to a specific disease state, ultimately allowing doctors to provide improved care for their patients.

This Dissertation

Following this introduction, in chapter 2, a short review of the technical details of PET will be presented; the scope and goals of this dissertation will be addressed in the final section of that chapter. Chapter 3 addresses the foundations of signal detection theory, and chapter 4 is dedicated to the science and implementation of mathematical observers. Chapter 5 describes a novel lesion detection experiment designed to quantify the gains in performance of various reconstruction methodologies. Finally, chapter 6 summarizes the research and the future prospects for the field of PET and possibilities for extending the study described in this dissertation.

Chapter 2

PET Physics

Fundamental Physics

PET scanners detect the energy released from positron annihilation. When an unstable isotope decays by beta plus (β^+) emission[2], a proton is transmuted into a neutron, emitting a positron and neutrino. As illustrated in Figure 5, the positron travels along a random path where it experiences inter-atomic coulomb interactions which influence its course and reduce its kinetic energy before encountering and annihilating with an electron. The mean free path of a positron in water is 0.6mm, but the actual path depends on the initial positron energy as well as surrounding tissue.

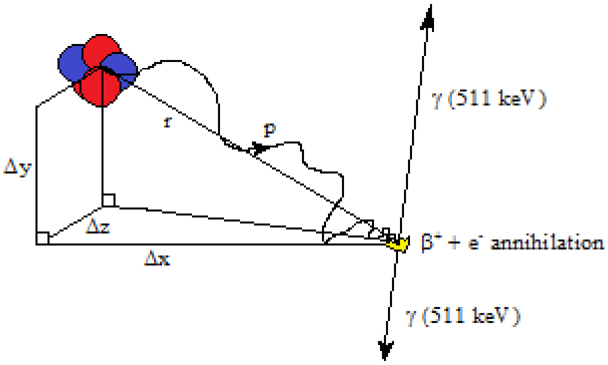


Figure 5. Positron emission and annihilation. Within the atomic nucleus, a proton (blue circles) transmutes into a neutron (red circles). This event releases a neutrino and a positron which then encounters a nearby electron, and the pair self-annihilate, releasing two 511keV photons in approximately opposite directions.

---

[2] In the case of fluorine-18, the branching fraction is 0.97 so there is a 3% chance that it will decay via electron capture. (Shapiro 2002)

Once the positron has lost enough kinetic energy, it will directly annihilate with a nearby electron or combine with it in a temporary state called positronium, where the positron and electron are in a bound state. In nearly all cases the positronium is formed in the singlet state, or para-positronium, one in which the two spins are antiparallel. The mean lifetime of this state is 124 picoseconds and can decay into any even number of photons, but the probability is drastically reduced for higher numbers than two, i.e. the branching ratio for decay into 4 photons is $1.439(2)\times10^{-6}$ (Karshenboim 2003). There is also a fractional probability that positronium will form in the triplet state, called ortho-positronium. This state has a mean lifetime of 142 nanoseconds and usually decays into three photons, necessary to conserve the charge parity of the positronium state.

Para-positronium essentially always decays into two 511 keV photons propagating in nearly opposite directions, thus satisfying the laws of conservation of mass and momentum.[3] The fundamental job of the PET tomograph is to detect these annihilation photons and identify coincident pairs arising from the same event.

Event Detection

PET detectors are sensitive to photons incident on the scintillator, although a certain fraction may pass through without interacting. Photons that stop within the scintillator are converted to light that is captured by the four PMTs. The distribution of light among the PMTs enables the position of the incident photon to be localized. The sum of the signals from the four PMTs is proportional to the energy deposited in the

---

[3] The relative alignment of the initial particle spins and the positron's kinetic energy at the time of annihilation may affect the number and direction of the emitted photons; however, these effects are usually considered negligible in practice. (Valk 2003)

crystal by the incident photon. The summed signal is required to lie within an energy window for the event to be accepted by the electronics. The width of the window is related to the energy resolution of the scintillators, and photons with a measured energy below the lower discriminator setting are generally scattered.

As mentioned previously, PET acquisition is based on detection of coincident pairs of photons, which it accomplishes through online electronic collimation. With every single photon detected, a timing window of a few nanoseconds is triggered. If a second detection is registered in an opposing detector within the timing window, it is recorded as a coincident, or prompt, event and attributed to the line of response (LOR) between the two detectors. Prompt events are relatively scarce, and the singles rate is generally 2 orders of magnitude higher than that of the prompts. Further confounding the problem, not all prompt events are true coincidences. Figure 6 shows the three types of prompt events that are recorded: trues, randoms, and scattered coincidences.

Trues are events in which both annihilation photons travel through the body and are detected without interaction in the surrounding medium. Randoms occur when two photons from different annihilations strike opposing detectors within the time window. The system incorrectly assigns the event to the LOR connecting the two detectors, one that may be far from the actual LOR. Scattered event detection happens when one (or both) of the photons scatters and is slightly deflected so that it maintains enough of its energy to be accepted within the energy window. For the present case, where a lower-level energy discriminator (LLD) of 435 keV is used, a photon can be deflected by 34 degrees and still satisfy the energy requirement, according to the Klein-Nishina formula.

Similar to the situation with randoms, this results in large errors because the system assigns the event to an incorrect LOR. If the scattered photon is deflected enough so that its final energy does not fall within the energy window, the event is rejected by the electronics. This loss of signal along the LOR is termed attenuation.

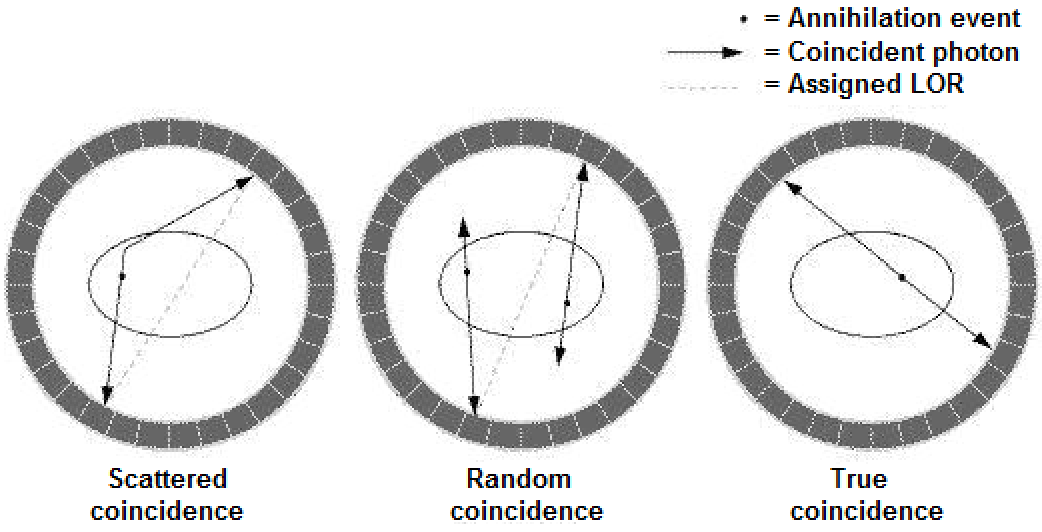


Figure 6. Diagram showing the 3 types of prompt detections: scatter, random, and true.

Corrections for attenuation, scatter, and randoms are necessary for the reconstruction process and are addressed further in a later section.

Acquisition Data

Historically, every prompt event from a scan was recorded in a long list. Originally, each prompt list entry included only the address of the respective LOR, but later, each prompt entry incorporated the relative arrival time of each coincident photon.

Additionally, timing tags (marking the time progression of the entire scan) and singles rates were added to the list. This storage mode is, not surprisingly, called list-mode. Storing the data in this way is beneficial for several reasons, particularly for dynamic scans where the list-mode data can be rebinned into appropriate time frames after the acquisition is completed. The rebinned frames are then input into the reconstruction algorithm.

As previously mentioned, the data acquired during a PET scan are recorded as integrals of the total activity lying along all LORs. The LORs are sorted into groups oriented at the same transaxial angle, forming parallel projections. For a given angle Φ, there exist a set number of parallel projections. A matrix comprising the projections for the complete set of N angles, sampled from 0 to π, is termed a sinogram. The first row of the sinogram matrix will be the projection for angle Φ, and the next row will be for $\Phi+(1/N)\pi$, etc. The matrix has dimensions equal to the number of sampled projections times the number of sampled angles. The term sinogram follows the fact that an off-center point source histogrammed this way will appear as a sinusoid.

The early PET systems were designed to acquire only projections at orientations orthogonal to the axis. These include the direct planes, which intersect crystals within the same detector ring, and the cross planes, interpolated from adjacent rings and interleaved between the direct planes. For a simplified example, a 2D PET system with two rings of crystals would a acquire a set of plane projections for each direct ring and one for the pseudo-plane between the rings, calculated from the events with detections in both rings.

This schema increases the axial sampling, producing 3 slices, and is directly extendable to any number *D* rings; the number of sampled slices is equal to 2*D*-1.

It is appropriate to introduce here the concept of detector span, which defines the acceptance range in axial ring difference of coincident photons. In the previous example, the span was 3, i.e. 0, which refers to the direct planes, and +/- 1, which contribute to the cross plane calculations. The case for span 7 is illustrated below in Figure 7.

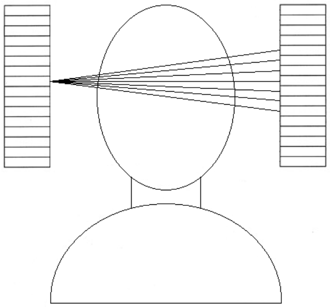


Figure 7. The detector elements in a system with span 7 only accept coincident data from those elements with locations ranging +/- 3 axially.

3D acquisition extends this concept to include oblique planes, or non-orthogonal planes that exist between crystal elements that lie in rings with difference greater than the span. This involves the removal of inter-ring septa and can increase the system sensitivity by a factor of around 5 (Bendriem and Townsend 1998; Tarantola, Zito et al. 2003). A 3D sinogram is an ordered collection of smaller sinograms corresponding to different detector difference groupings, called segments.

Figure 8 below expands on our earlier example. A coincident event registered in two detectors, within the 7 element span (centered on the direct plane), will belong to segment 0. The second segment (+1) sinogram will include events from elements where the difference is greater than or equal to 4 but less than 11. The next segment (not shown below) will be composed of events from detectors whose difference is less than or equal to -4 but greater than -11. The final segment shown below includes the events from elements where the difference is greater than or equal to 11.

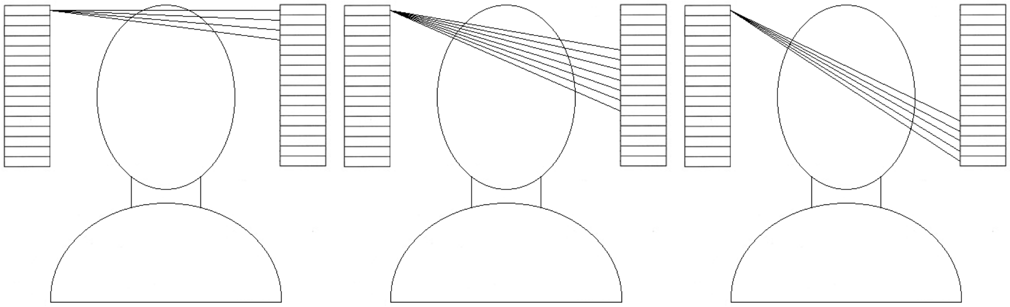


Figure 8. Representation of the different segments acquired in 3D acquisition.

A 3D sinogram will follow this trend including the associated sinograms from each segment.

Data Corrections

Radioactive decay is a random process which follows a Poisson distribution, i.e. there is a discrete probability that an event takes place in a given time interval, a probability that is temporally and spatially independent from other events. Most reconstruction algorithms model this Poisson process and treat the data as if they are free of noise or error, but in reality, the PET acquisition is an intrinsically noisy process. There are many physical effects that can corrupt the integrity of the data, including detector normalization, photon attenuation, and randoms and scatter registration. These errors degrade image quality by adding noise to the PET data and must typically be corrected prior to (or incorporated within) the reconstruction, for accurate image quantification.

Randoms

Continuing from the previous section, random coincidences are recorded when two photons from different events hit two detectors within the timing window. These events contain no useful information about the tracer distribution and, because they are random in nature, add smooth background noise to the data. The random coincidence rate is related to the size of the timing window and the singles rates on the detectors.

A few methods have been proposed to model the randoms distribution, including detector singles-based calculations and tail-fitted, Gaussian estimations (Valk 2003). Currently however, the most commonly implemented technique is the delayed channel or delayed-window subtraction (DWS) method. Here, a timing window is delayed (a few times the timing window length) to acquire a duplicate data stream in parallel to the prompts. The delayed timing window guarantees that any recorded coincident photon pair does *not* belong to the same annihilation event. These delayed data provide accurate representation of the randoms rate within the prompts window and can be stored independently, allowing post-processing the randoms data to reduce noise.

Scatter

As previously mentioned, scattered events occur when at least one photon of the coincident pair is Compton scattered, deflecting it before hitting the detector ring. Scattering decreases the energy of a photon, so this effect is addressed by acquiring only the photons lying within the photo-peak energies, shown in Figure 9. This is accomplished with an appropriate energy-acceptance window (typically 450-650 keV for LSO-based scanners). However, a photon can scatter through a relatively large angle and still have enough energy to be accepted. Thus, many scattered events are not rejected by the energy discriminator.

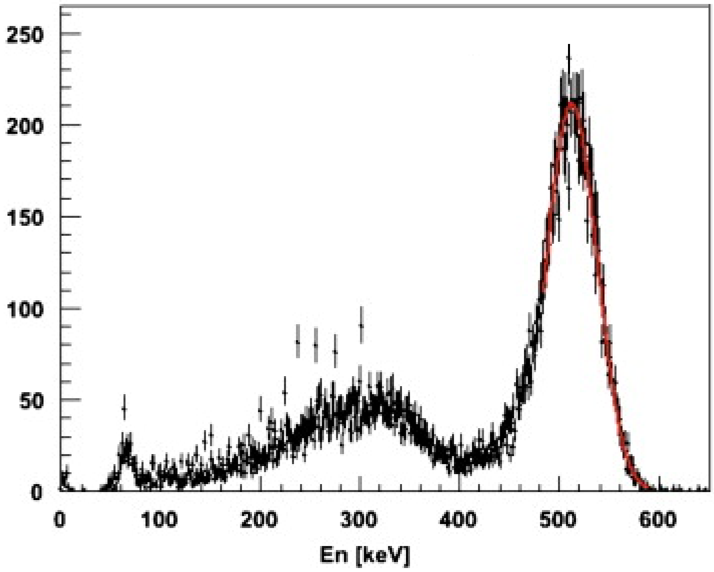


Figure 9. Typical energy distribution of scintillation detected energies.

Scatter is an important problem in 3D PET, as in a clinical scan, 30-50% of the data may be from scattered photon events (Cherry, Dahlbom et al. 1991).

Modern PET scanners acquire data almost exclusively in 3D mode where the added sensitivity to noise, and especially scattered events, requires more sophisticated correction methods. Empirical schemes have been used that measure auxiliary data, such as the coincidence rates at energies below 511keV, under the assumption that these lower energy photons must be the result of scattering. By employing multiple energy windows, these auxiliary data are scaled and subtracted from the data in the photo-peak window (Grootoonk, Spinks et al. 1996). Scatter correction based on multiple energy measurements has the advantage of accounting for scatter arising from activity outside the field of view. However, the auxiliary data are noisy and their processing requires more computing power. Another technique is to estimate the scatter distribution from prompt counts registered at places known to contain no activity. Scatter has a low frequency, broad distribution with tails that extend beyond the boundaries of the object. These tails can be used to fit a smooth Gaussian function to the scatter distribution (Stearns 1995). This model performs reasonably well in brain studies, but can lead to errors in whole-body scans where the thorax occupies a larger portion of the FOV, resulting in relatively small scatter tails to fit.

Some approaches used in 2D PET (with inter-ring septa) are convolution-based integral transformations of the projections (Bergström, Eriksson et al. 1983). In this method, the scatter distribution is estimated through the convolution of a spatially dependent scatter “kernel” with the linear projections of the prompt photo-peak data. The resulting distribution profiles are then subtracted from the projections to yield scatter-free estimates of the data. It works well in regions of relatively homogeneous density like the brain. This method was developed for 2D PET but has been extended to 3D systems (Bailey and Meikle 1994).

Modern 3D PET algorithms, however, most commonly use theoretical simulation techniques to correct for scatter. They are arguably the most accurate scatter corrections, in that they attempt to realistically simulate the scattering process. Scatter simulations model the distribution based on the underlying physics of Compton scattering. With an initial estimate of the emitter distribution and a map of the linear attenuation coefficients of the attenuating material, Klein-Nishina calculations are used to accurately compute the scatter distribution (Ollinger 1996). This map can be obtained by scaling the transmission images from the CT (Beyer, Kinahan et al. 1994).

Monte Carlo simulation is widely used to model scatter and provides the most complete realization of the physical effects associated with it. This technique is used for both for the correction and for the evaluation of the correction, but it is currently too slow to be used clinically and is limited to research applications. Instead of accounting for all scattering interactions, the model can be simplified by considering only single scatter events, i.e. events where just one photon of the pair is scattered only once (Watson, Newport et al. 1996). This simplification is reasonable since it has been shown that 75-80% of scattered coincidences are due to these single scatter events (Barney, Rogers et al. 1991). It is further justified by the fact that multiple scatters, unlike single scatter, are not highly correlated with the underlying emission distribution.

The single scatter simulation (SSS) calculation relies on a volume integral of a scattering kernel over the body, using the relationship between an initial estimate of the activity distribution volume,[4] the reconstructed attenuation volume, and Compton scattering cross-sections calculated from the Klein-Nishina formula. The scatter contribution is calculated for every LOR across all possible scattering points in the FOV. A faster method has been proposed that is better suited to clinical application (Watson 2000). With this method, fewer LORs are used in the calculation and continuity is replaced by discrete sampling of the volume data. The sample points are relative to the transaxial FOV (not the patient) and, since the scatter distribution is relatively smooth, the sampling can be fairly coarse (around 2 cm) without much loss of accuracy. This coarse-scatter sinogram is then interpolated to account for the missing data. Speed is further improved by reusing computed ray sums through the object since scatter calculations in multiple LORs may involve the same photon travel paths (Mumcuoglu, Leahy et al. 1996). Accuracy is further improved by iterating the scatter calculation, i.e. with each iteration, the previously corrected emission is used in the calculation.

Attenuation

Compton scattering results in 511keV photon attenuation along each LOR. It is directly related to the amount of matter (tissue) along the LOR, so attenuation is a bigger problem in larger patients. Illustrated in Figure 10, the loss of count data results in reduced image contrast and detail for internal features, especially those lying deep inside the subject.

---

[4] Different methods have been employed for initial estimates of the emission volume distribution; Watson et al. used 3D reconstructions of the projection data, which include scatter. (Watson, Newport et al. 1996)

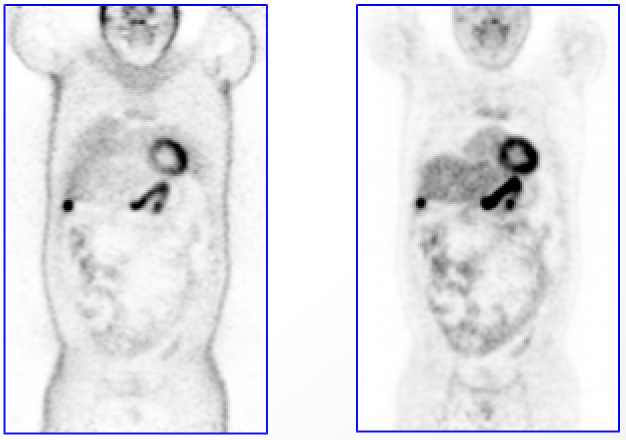


Figure 10. Attenuation reduces information about internal structure; the image reconstructed with corrected data is shown on the right

With PET, the photon attenuation for each LOR is independent of the location of the annihilation. A correction factor is therefore assigned to each LOR, proportional to the total attenuation along that line. The coefficients are termed attenuation correction factors (ACF) and stored in a matrix similar to the emission sinogram. In the simplest sense, attenuation correction is the multiplication of the PET projection data with the ACF sinogram.

Before the introduction of hybrid PET/CT, ACF sinograms were generated using a 511keV photon transmission scan of the subject prior to the emission measurement. The ACFs are calculated as the ratio of the transmission scan to a blank scan - a scan acquired with no patient in the field-of-view. The blank and transmission scans were acquired by rotating a line source of positron activity (usually ^{68}Ge) around the patient to provide information about tissue density. Sinogram windowing and data from the subsequent emission measurement were used to remove most of the emission counts from the transmission data. Furthermore, FDG studies have relatively low count rates, and there is little increase in noise in the correction due to emission activity (Carson, Daube-Witherspoon et al. 1988). These methods produced accurate ACF maps, but required increased patient time in the scanner.

With the combination of PET and CT, it is possible to generate ACFs using the CT transmission scan. It is no longer necessary to include PET transmission sources with the scanners, eliminating their initial costs as well as those associated with the periodic replacement of decaying sources. The high flux of X-rays leads to lower statistical noise in the ACF measurements and shorter transmission scan times (seconds rather than minutes).

The challenge of using the CT transmission data for the PET attenuation correction arises from the fact that the photons are of different energies. For annihilation photons at 511keV, the probability of photoelectric absorption is nearly zero, and thus the attenuation results almost exclusively from Compton scattering. The X-ray photons are less energetic, around 70-80keV (Townsend 2008), and the effect of photoelectric absorption is also significant in this range. Thus, the attenuation data from the X-ray CT must be scaled prior to use in the PET corrections. This problem is not straightforward, since the scaling depends on the photon energy and the tissue density in the FOV. Linear scaling of the CT to PET attenuation map yields a poor estimate, especially in high-density regions like bone (Kinahan, Townsend et al. 1998). For this reason, density segmentation of different tissue regions has been used successfully, but this method produces errors in regions where there are variations in density, such as the lung (Beyer, Kinahan et al. 1994).

CT-based PET attenuation factors are now routinely computed through a combination of both, segmentation and linear scaling. The relationship of attenuation coefficients at CT and PET energies is roughly linear for all soft tissue (Hubbell and Standards 1969), so current methods segment regions below 80 Hounsfield units[5] as soft tissue, and above 80 as bone tissue. The CT-to-PET ACF scaling is 2.26 and 1.90 (Shreve and Townsend 2008), for each respective region; the scaling is shown in Figure 11.

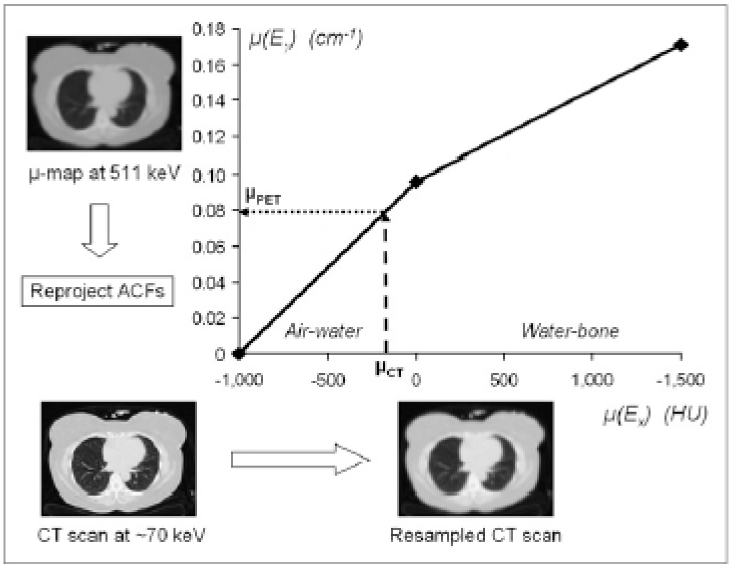


Figure 11. Bilinear scaling function used to convert CT numbers (HUs) to linear attenuation values at 511 keV. The ACFs are generated by reprojecting the scaled map.

---

[5] Houndsfield units are used in X-ray CT to represent the electron density of biological tissue. The values are calibrated so that water is zero, with more and less dense material lying in the positive and negative ranges, respectively.

This bi-linear scaling has shown excellent results that perform superior to 511keV transmission methods, with lower noise (Kinahan, Townsend et al. 1998) .

Normalization

Reconstruction algorithms generally assume the same sensitivity for all LORs, but this is not in general the case. Differences in crystal sensitivity and photomultiplier tube gains, as well as scanner geometry and detector block gaps, contribute to varying sensitivities for the different LORs. As a result, quality control must be performed regularly to ensure that the PET scanner is operating optimally and that the sensitivity profiles are known for all LORs. A uniform distribution of known activity is placed in the field of view, and the measured sensitivity profiles are stored for use in the normalization correction of routine clinical scans.

Shown below in Figure 12 is an example of a sinogram before and after corrections are applied for attenuation, randoms, scatter, and normalization.

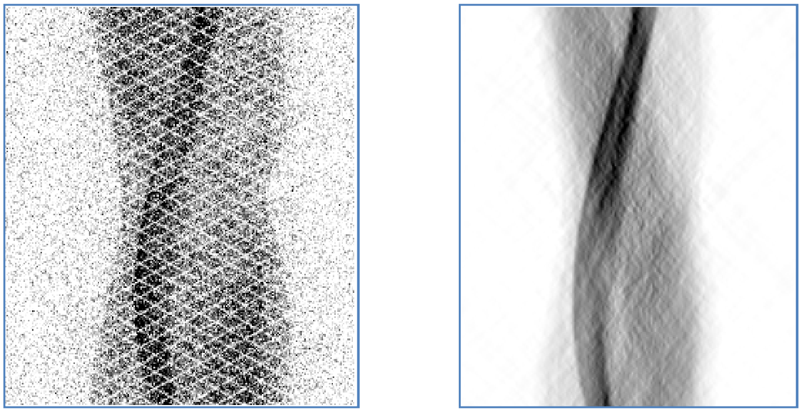


Figure 12. Raw sinogram (left) and the same sinogram corrected for attenuation, randoms, scatter, and normalization.

Reconstruction

Over the course of the scan, millions of counts are recorded, distributed among the various LORs within the FOV. For every angle, the LORs are sorted into sets of parallel projections, where each projection bin is the 1-dimensional integral of the measured activity along the line. The basic task of any tomographic reconstruction algorithm is to estimate the spatial distribution that produced the measured projections.

Direct Reconstruction

The projections are expressed mathematically as Radon transforms of the data. The reconstruction attempts to form an image through the inverse process. However, if the projections are simply propagated back along their respective LORs, low frequency "smearing" effects dominate the resulting image. Figure 13 demonstrates this effect.

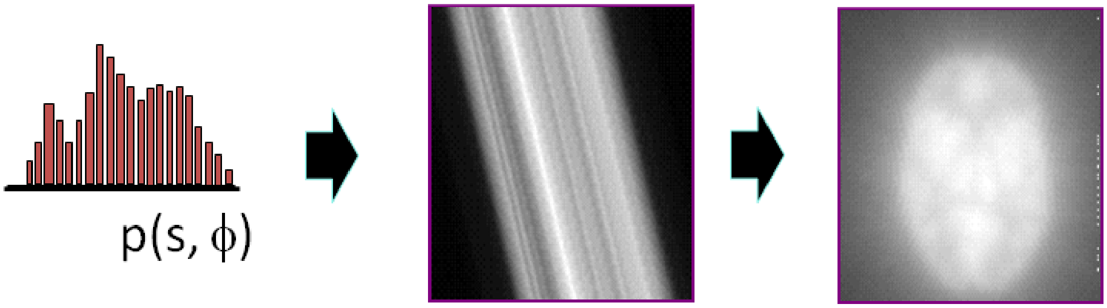


Figure 13. Blurring effect of backprojection

The effects of the backprojection operation cause the reconstructed image to appear blurry because every point in the original distribution is more closely sampled at small radial distances. In fact, the image produced from the backprojection process $f'(r)$ is represented mathematically as the convolution of the original distribution $f(r)$ with a radial blurring kernel $1/|r|$. The commutative nature of the projection integral permits the blurring to be pre-corrected by filtering the projections before backprojecting them. A frequency ramp filter is used to introduce negative values in the projections and suppresses the low frequency contributions in the final image. This is appropriately called filtered backprojection (FBP) and is illustrated in Figure 14.

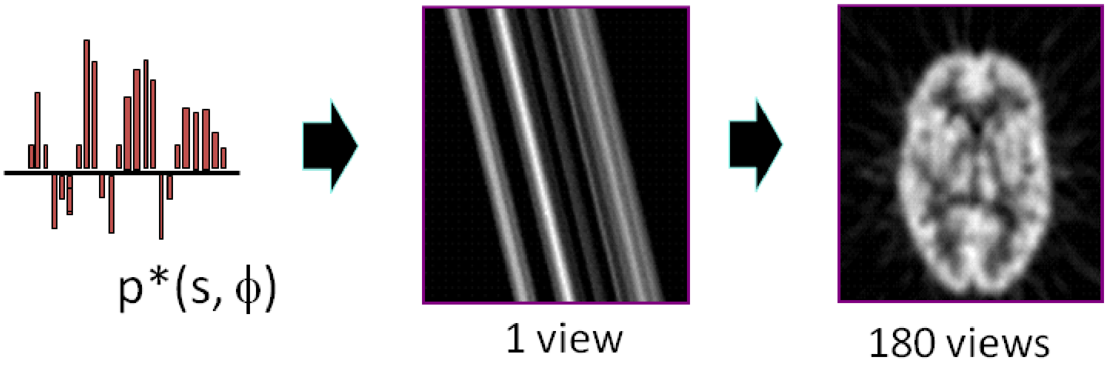


Figure 14. Filtered backprojection process

The ramp filter amplifies the high frequency components in the back projection. The filtering process is based on the central slice theorem, which is also the foundation for direct inverse Fourier transformation (DIFT). In this scheme, the projections in frequency space are interpolated and used to create the Fourier image which is then directly transformed back into the real image.

PET is an intrinsically noisy process. Statistical noise in Fourier space has high frequency components, and the ramp-filtering process amplifies this noise. Due to this, other windows are introduced to reduce this unwanted amplification. Even so, the quality of images generated through direct methods is generally inferior to that of non-linear algorithms.

Iterative Reconstruction

Most current PET reconstruction algorithms do not produce images by directly inverting the projection data. Instead, they are based on sequential, iterative estimates that converge to an image best representing the original object. These algorithms are termed expectation maximization (EM) and are well suited to low-count emission tomography because they are less sensitive to noise in the data.

Iterative reconstruction involves the sequential forward and backprojections of a simulated image in the attempt to match the measured distribution as accurately as possible. Using this technique, the image is approximated, forward projected, and then compared with the actual measured projection data. The difference in the projections is used to generate a correction matrix which is applied to the estimated image. This is done multiple times until the algorithm arrives at an acceptable image. The basic EM algorithm takes the form

$$X_j^{k+1} = X_j^k \cdot \frac{1}{\sum_i p_{i,j}} \sum_i \left[p_{i,j} \cdot \frac{Y_i}{\sum_j p_{i,j} \cdot X_j^k} \right] \tag{2.1}$$

The estimate image X changes with each iteration k. Integral to this method is the system matrix p which models the physical characteristics of the scanning process. It contains the discrete probabilities that each projection bin i contributes to each image pixel j. Ideally, the system matrix accounts for all physical effects of scanner geometry and performance. Attenuation-weighted (AW) algorithms, commonly used in clinical practice, incorporate the attenuation map from the CT scan into the system matrix, and the forward projector is better able to realistically simulate the acquisition process. This increases the accuracy of the physical model unique to the individual scan.

Iterative algorithms did not find their place in the clinical setting until the last decade because the computational requirements of the iterative approach made the reconstruction process very slow. The projections are calculated over many angles, typically 128, 168 or 336, each taking as long as a complete FBP reconstruction. Figure 15 shows the qualitative improvement of iterative reconstruction over filtered backprojection, both reconstructions being of the same projection data.

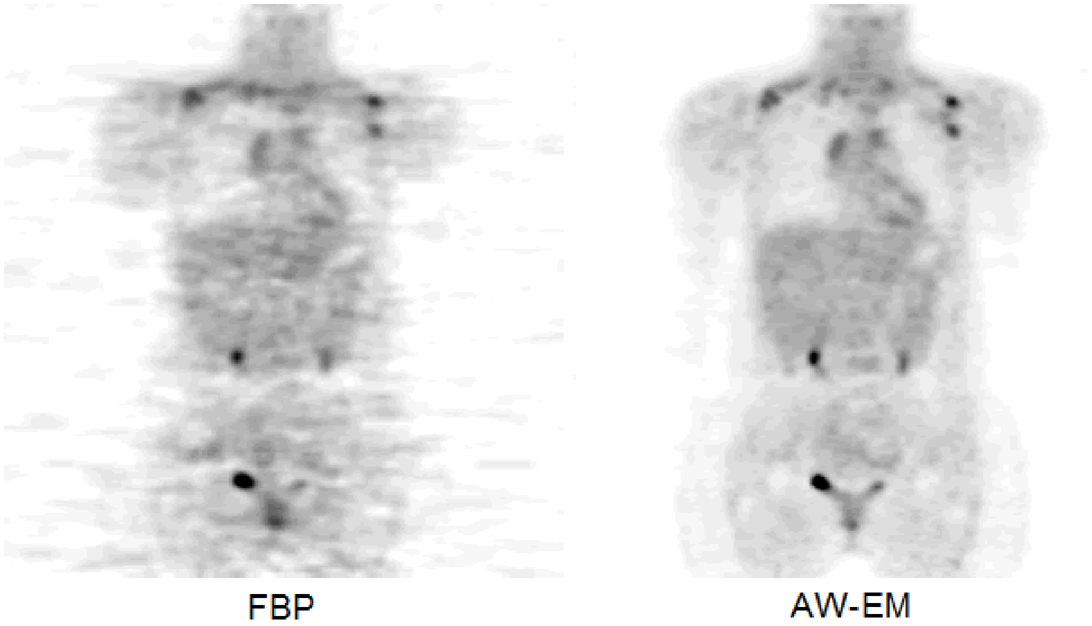


Figure 15. Expectation maximization (right) is less sensitive to noise relative to filtered backprojection (left) and produces better image quality in low-count conditions.

Scanner Performance

Image quality has drastically improved since the beginning of PET, due largely to hardware improvements within the tomographs. Advances in scintillating materials, larger numbers of detectors, smaller crystal elements, and faster electronics all improve the photon detection system. The performance of an individual scanner can be defined in terms of its noise equivalent count rate (NECR).

$$NEC \equiv \frac{T^2}{T + R + S} \tag{2.2}$$

The NECR is the ratio of true detections to the total and is dependent upon the activity; it is a useful metric for quantification comparisons across multiple systems.

Recent Developments

Gantry design, electronics, and image processing have advanced PET greatly over the past few decades. One of the most important developments was the addition of a spiral CT to the PET scanner, giving us the PET/CT. As previously mentioned, the CT plays a huge role in modern image acquisition; it decreases scan time while improving image quality.

Another important development was the integration of LSO crystal elements into the PET detector rings. Invented in the early 90's, LSO (cerium-doped lutetium oxyorthosilicate) has higher light output, and better energy resolution than BGO.(Melcher and Schweitzer 1992) In addition, the decay time for LSO is nearly ten times smaller, allowing the capacity for time-of-flight measurements.

Time-of-Flight

Conventional PET scanners assign annihilation events to LORs then record the acquisition data as projections across the entire FOV. Time-of-Flight (TOF) refers to the additional measurement of the coincident photons' relative time of detection. Using this information, the system can approximate the location of the annihilation along the assigned LOR.

The idea of using the timing of coincident events to improve PET images, shown in Figure 16, has been around since the beginning of PET (Allemand, Gresset et al. 1980;

Mullani, Markham et al. 1980; Tomitani and Tanaka 1980). However the performance of TOF detectors limited its use to research purposes; with LSO, the practical application of photon timing measurements became possible in clinical PET.

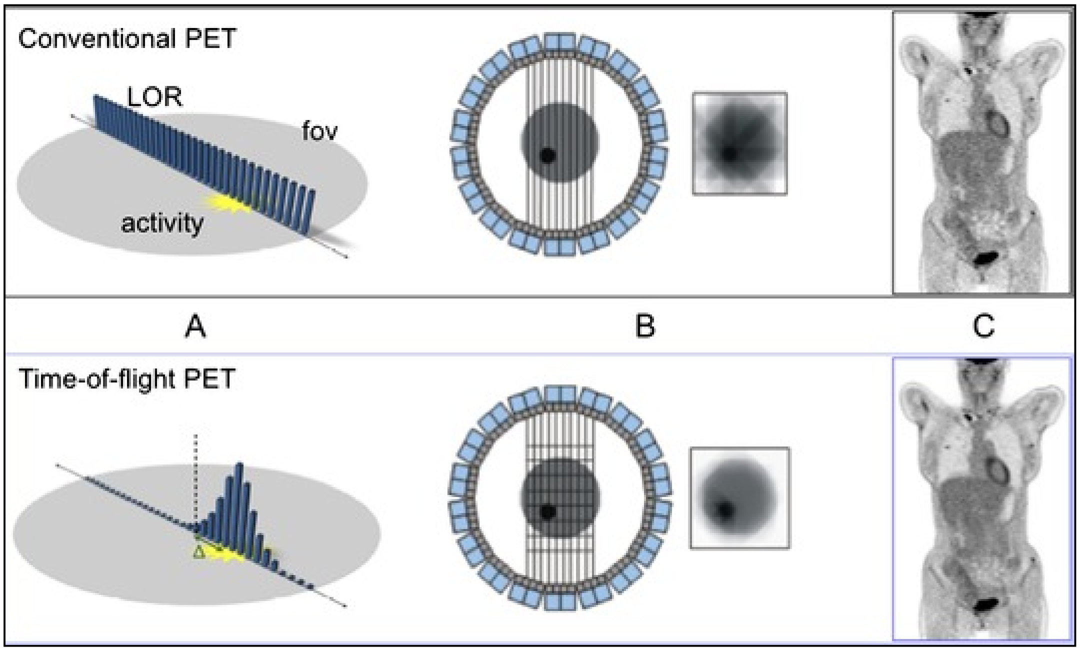


Figure 16. Time of flight information reduces the noise in the reconstruction and improves image quality.

The geometry of the TOF-PET gantry is the same as that of conventional PET, with multiple adjacent rings of detectors surrounding the patient, but the TOF systems are capable of measuring the small timing differences within the photon pairs. This information is used to locate the approximate position of this event along the line. So instead of granting equal probability of the event along the line, the system calculates the probabilistic location of the event, reducing the propagation of noise along the LOR.

This reduction of statistical noise can be interpreted as increased sensitivity of the detection process.

The precision of this localization, namely how accurately the detection times can be measured, depends on the timing characteristic of the detection system, including the crystal properties and electronics performance. If we place a point source at the center of a detector ring, it will be equidistant from each detector element. Thus, the annihilation photons should reach any detector at the same time. However, due to the uncertainty in the time measurement, the total number of measured events does not form a sharp peak at the same value but rather forms a Gaussian distribution. The accepted timing resolution rating of a system is the full-width at half-max of this spread.[6]

In FBP reconstructions, the gain in signal to noise with added TOF information comes from the ratio of the diameter of the object to be imaged to the timing resolution of the system (Casey 2008). In the clinical setting, TOF information is more helpful in larger patients, and smaller time resolutions allow better, more certain measurements. In iterative reconstruction methods, the addition of TOF information allows the algorithm to achieve higher contrast recovery with less noise in fewer iterations (Conti, Bendriem et al. 2005; Karp, Surti et al. 2008; Lois, Jakoby et al. 2010) .

TOF-PET promises the possibility of higher quality medical images, but due to only relatively recent developments, TOF-PET has not yet become fully integrated into

---

[6] Theoretically, a system with perfect timing resolution would be able to generate an accurate image using only the spatial and temporal detection event information, eliminating the reconstruction process altogether. However a timing resolution less than 30ps would be required to localize the event to a 4mm pixel, this is nearly 20 times less than the best performance achieved today. The intrinsic spatial limit of PET is set by the mean free path of the positron before annihilation. If we wish to achieve resolution on this order, a timing resolution of 4ps is required (based on the .6mm mean free path of an ^{18}F positron in tissue).

mainstream clinical applications. Therefore, its clinical benefit has not been extensively investigated.

Point Spread Function

The spatial resolution across the transaxial image plane is not isotropic; rather it degrades with radial distance from the center. Shown below in Figure 17, the circular geometry of modern-day PET scanners causes many annihilation photons in LORs far from the center to strike two detectors at highly oblique angles. This leads to depth-of-interaction effects within the crystal block and event misregistration.

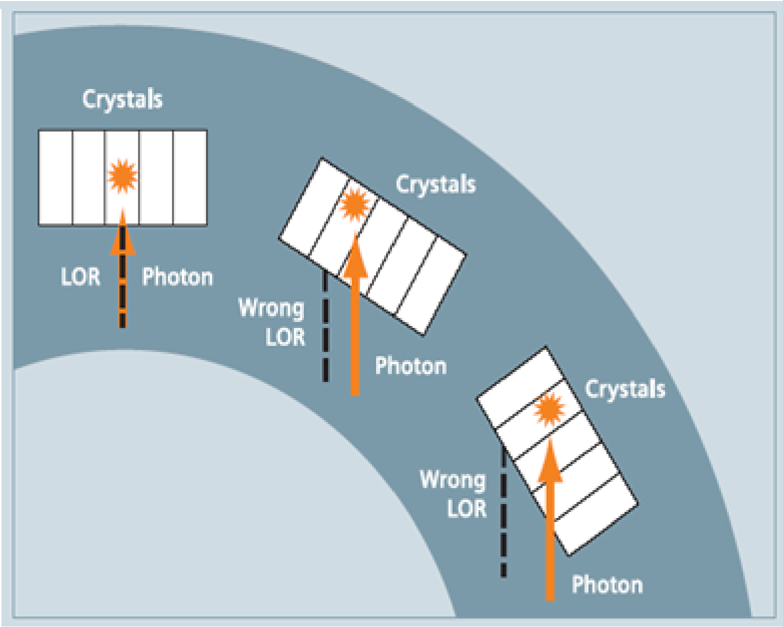


Figure 17. The circular geometry of PET scanners introduces registration errors for photons striking the detectors at highly oblique angles, blurring the system point response, especially for LORs at large radial distances.

This parallax effect and inter-crystal scattering cause the point response of the detection system to become blurred.

This phenomenon has been incorporated into EM reconstructions. Point sources are independently acquired at different radial locations, and the spatial response, unique to the scanner, is measured and stored (Panin, Kehren et al. 2006). As expected, the shape of the PSF model broadens, for each radial bin within the sinogram, as the distance from the center increases. This information is then incorporated in the system matrix so that this effect is modeled in the forward and backward projections of the iterative algorithm.

This model reduces high frequency noise within the image and improves spatial resolution, especially at large radial distances. Both of these effects occur because the point spread functions act like matched filters, enhancing data that matches the model and penalizing the noisy data that does not (Casey 2007); incorporation of the PSF model complicates the projection steps of iterative algorithms and results in delayed image convergence.

Research Goals

PET imaging has undergone considerable development since the 1980s that has led to the integration into clinical practice. These include improvements to the both the system hardware and software. Ongoing work is essential to characterize the advances in performance from the introduction of new methods and technologies. In many cases, it is also necessary to translate this to clinical applications, and to assess the impact that these new advances have on overall patient care.

This research focuses on the incorporation of point spread function modeling and time-of-flight information within PET reconstruction. Throughout the course of this work several questions are addressed, the most notable being:

1. What are the optimal parameters of the iterative algorithm and processing schemes when the point spread function and time-of-flight models are integrated in clinical reconstruction?
2. How does the optimization of these models improve clinical image quality, in terms of observer signal to noise?
3. Do these models improve detectability of small focal lesions in clinical images, and can these improvements be quantified?
4. Which numerical observer performs best in this task?
5. Will the data from experienced radiologists validate those from mathematical observers, and which model observer agrees most closely with humans?

Chapter 3

Signal Detection Theory

Receiver Operating Characteristics

An observer in a classification task is understood to arrive at a decision by making a perceptual measurement of an object; this measurement is then compared to some threshold, to determine to which underlying hypothesis the test object belongs. It can be argued from the central-limit theorem that the decision probability functions of the measurement variable will be normally distributed, over observer threshold criteria.

Observer performance in a classification or detection task cannot generally be characterized by accuracy; the prevalence of one underlying hypothesis over any other will bias this measure. Instead, performance is defined in terms of the rate of correct and incorrect decisions. For example, in a signal detection task, an observer will make a given number of detections at any decision threshold, some true and some false. If the threshold is too low, many non-target signals will be detected giving rise to a high false positive rate. Conversely, if the threshold is too high, it will eliminate the detection of some false positive decisions, but at the cost of a lower rate of detection of the true positives.

A receiver operating characteristic (ROC) curve is a plot of the true positive rate against the false positive rate, over a full range of threshold values. The ROC curve completely defines the observer performance in a binary classification task. The area under this curve (AUC) is a well-accepted figure-of-merit for a given observer system

(Barrett and Myers 2004). A value of 1 means the system performed perfectly and value of 0.5 means the observer performed no better than blind guessing.

Signal detection experiments measure observer ROC statistics using samples that are known to have come from both classes, in our case, abnormal (target) or normal (non-target) images. If the probability distribution of the decisions in the non-target locations is $g(X)$, the probability of a false positive decision is given by

$$P_0(Z) = \int_Z^{\infty} g(X)dX \tag{3.1}$$

or equivalently

$$P_0(Z) = 1 - G(Z) = 1 - \int_{-\infty}^{Z} g(X)dX \tag{3.2}$$

Here, Z is the observer’s decision threshold and $G(Z)$ is the probability that the measurement at the non-target site is less than Z. This model assumes that the presence (or absence) of lesions has no effect on the observers perceptual measurement, ie. a false positive measurement on a non-target location would have exactly the same measurement on the corresponding target image. Here, the decision would only be changed if the measurement of the actual target location exceeded that of the non-target one. Thus, the false positive rate is defined by the decisions within the non-target locations and is completely characterized using only the non-target images.

Similarly, we define $f(X)$ as the decision probability function in the distribution of target locations. The positive decision rate in the set of target *images* is not completely defined by $f(X)$ however, since some non-target locations may yield higher perceptual measurements than those of the actual targets.

We look at the simple case where there is only one type of lesion present in each target image, and the observer reports if it is present or not. The observer reports a positive decision if the perceptual measurement at the lesion site *or* at any other (non-target) site is greater than the observer threshold. The probability of a positive decision in the target class is given by

$$P_1(Z) = 1 - \mathrm{H}(Z) = 1 - \int_{-\infty}^{Z} h(X)dX \qquad (3.3)$$

where $\mathrm{H}(Z) = \mathrm{F}(Z)\mathrm{G}(Z)$ is the combined cumulative probability of the choosing the maximum value of the target or non-target site in target images. The distribution $h(Z) = f(Z)\mathrm{G}(Z) + g(Z)\mathrm{F}(Z)$ is the joint density function. Figure 18 and 19 illustrate this.

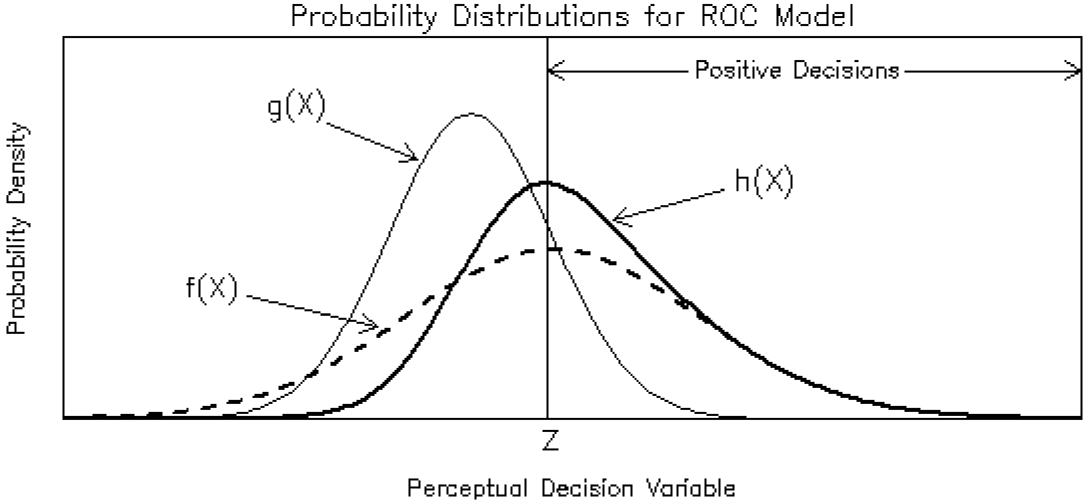


Figure 18. Probability distributions for the perceptual variables assumed to determine the ROC curve

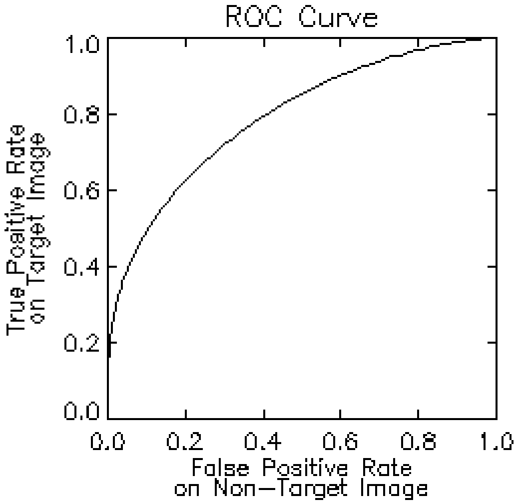


Figure 19. ROC curve generated from the distributions in Figure 18

The detectability index is directly related to the AUC. Thus, the SNR can be determined for a human observer, without knowledge of the specific test statistics.

Location Receiver Operating Characteristics

This is provides a good segue to the extension of the ROC to include localization performance, namely the localization receiver operating characteristic (LROC) formulation. The ROC model neglects localization information and treats all decisions as binary image ratings. The performance of correct lesion localization can also be represented as a function of false positives, similar to the ROC, but the curve does not go to unity (except for the slim chance of a 100% rate of correct localization). The distribution of observer decisions in the target images comprises those from target and non-target locations.

The target-image distribution in the ROC formulism consists of two mutually exclusive groups, correctly and incorrectly localized lesions. The probability of correctly localizing a target is the joint probability that the target location yields a perceptual measurement greater than Z and that this measurement is greater than that of all other (non-target) locations. It is given by

$$P_{CL}(Z) = \int_{Z}^{\infty} f(X)G(X)dX \tag{3.4}$$

Similarly, the probability of incorrectly localizing a target in the target images is

$$P_{IL}(Z) = \int_{Z}^{\infty} g(X)F(X)dX \tag{3.5}$$

The respective LROC curves are seen in Figure 20 and 21.

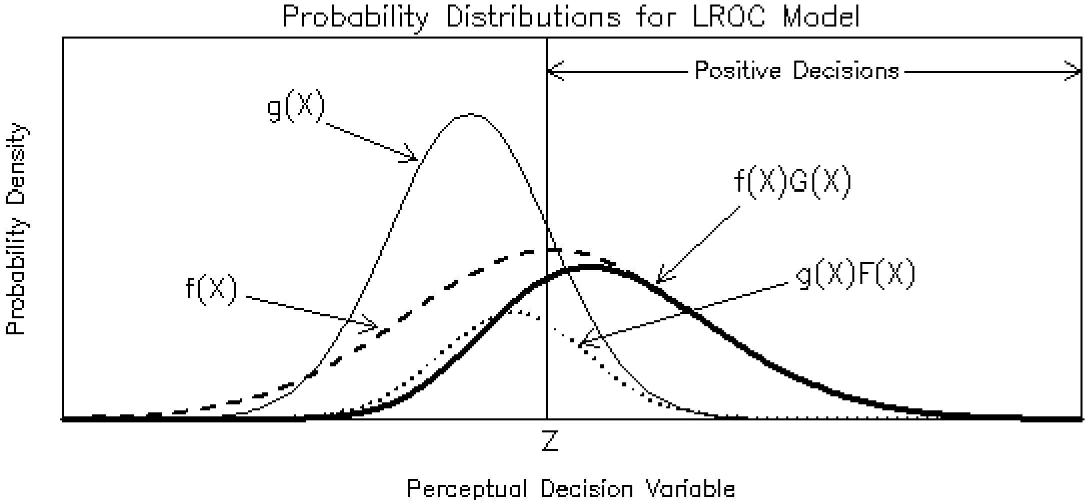


Figure 20. Probability distributions for the LROC model

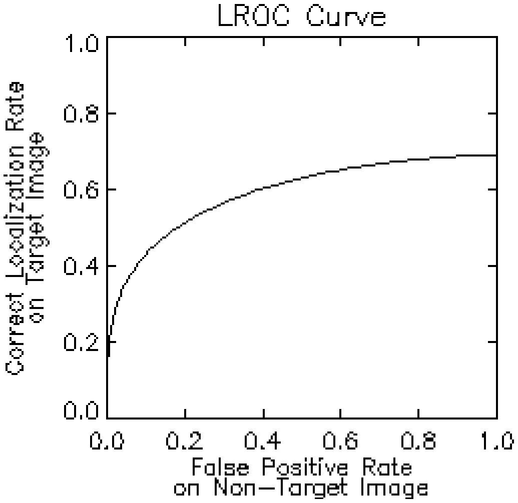


Figure 21. LROC curve from the distributions in Figure 20

Other derivatives of the ROC model include free-response receiver operating characteristic (FROC) and alternative free-response receiver operating characteristic (AFROC). Under the FROC paradigm, the images are not limited to only one known lesion. The observer has no knowledge of the number of lesions in each image, but chooses and rates multiple locations. The FROC curve is a plot of the correct localization rate versus the rate of incorrect localizations per image; it can extend indefinitely to the right. The AFROC is also a free response task, and like the FROC curve, its ordinate is the fraction of correctly localized lesions. However the abscissa is the false positive fraction, and like (L)ROC, the curves are constrained to the unit square.

Free response tasks are statistically superior to those of ROC and LROC, but some data suggest that human observers may violate “stationary performance” assumptions when making multiple report interpretations of images (Swensson 1996).

Confidence Rating

To completely characterize observer performance, each flavor of the ROC formalism needs decision information across a full range of threshold values. In a signal detection experiment, this is surveyed through the use of a decision rating scale. Each perceptual measurement is accompanied by a rating that reflects the observer’s confidence in the decision. Under the assumption that the observer’s decision criteria remain static throughout the experiment, the confidence rating is a surrogate for observer threshold. For example, say an observer detects a lesion at a given location with a confidence of 50%. At another location (in that image or another), he/she identifies another lesion with 90% certainty. It is reasonable to assume that the lesion rated at 90% would also have been detected with the more relaxed criteria of the 50% lesion, but not vice versa. It follows that every lesion rated at a certain confidence level would also be detected by the perceptual standards of any of the lower levels. The example below illustrates how this method is implemented in the ROC paradigm.

Table 3. Sample of confidence-rated decision choices for a binary task

Rating	***Non-target Distribution***	***Target Distribution***
0-20%	9	1
20-40%	6	4
40-60%	4	7
60-80%	1	6
80-100%	0	2

These data are used to fit a normal kernel density estimation of the distribution functions. The resulting cumulative and probability density functions are shown below in Figure 22 with the respective ROC curve.

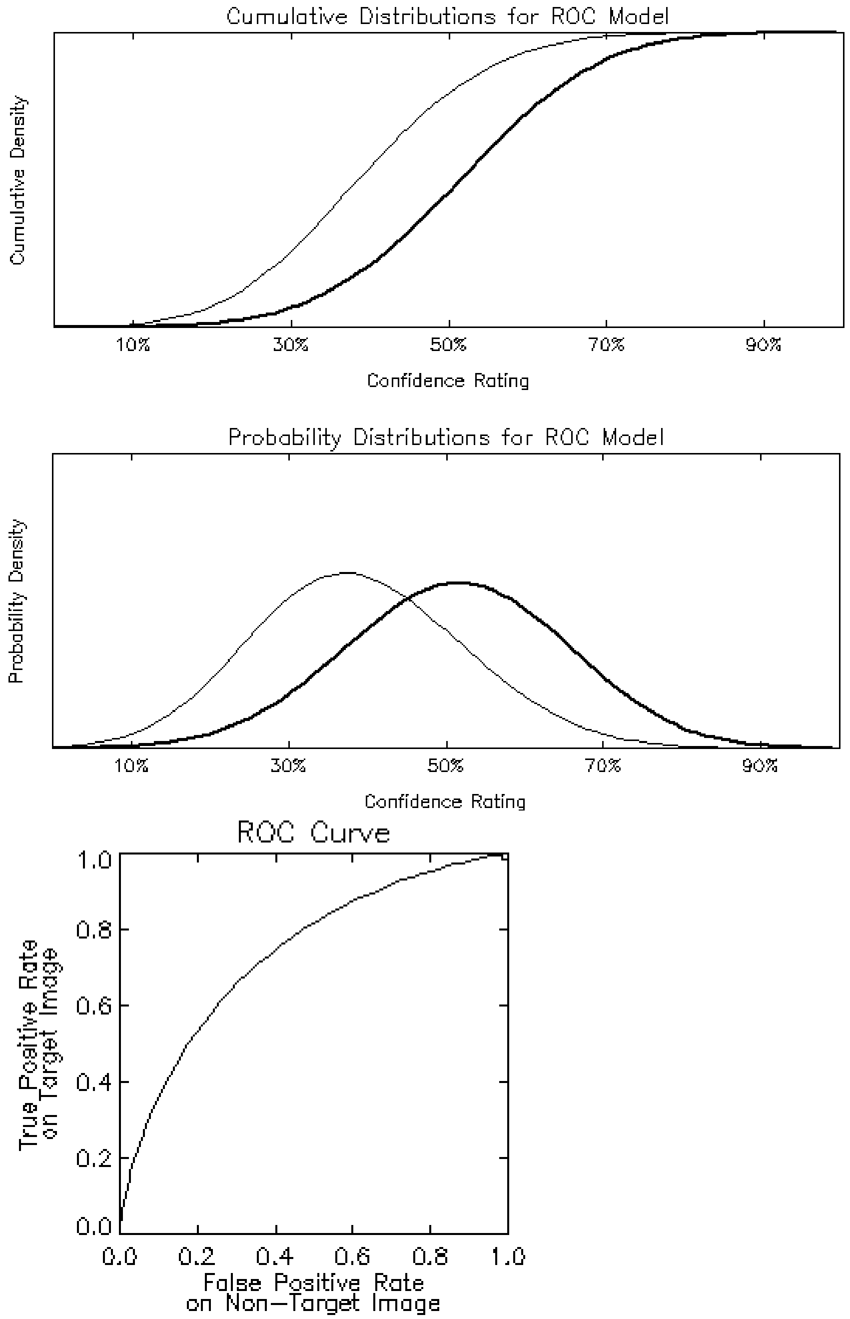


Figure 22. Cumulative and probability density functions, along with the resulting ROC curve fit from the sample confidence-rated data in Table 3

Parametric Fitting

The distribution equations defined above constrain both the ROC and LROC curves to be monotonic, i.e. each point on the curve defines the lower bound of operating performance at that threshold level. However, this method of fitting two data sets independently was designed for less complicated analyses, and it may not always impose monotony in the ROC curve (Kadrmas, Casey et al. 2009). This results in nonsensical performance curves, which may be negative or exhibit an upward hook-like inflection.

An alternative parametric ROC fitting technique has been developed (Swensson 1996) which mutually relates the two probability distribution functions, using only the relative means and variances of the distributions. The non-target decision distribution $g(X)$ is arbitrarily scaled to standard (zero mean and unit standard deviation) form, represented by the Gaussian curve

$$g(X) = \phi(X) \equiv \frac{1}{\sqrt{2\pi}} e^{-X^2/2} \tag{3.6}$$

The target distribution is then defined as

$$f(X) \equiv \left(\frac{1}{\sigma}\right) \phi\left(\frac{X-\mu}{\sigma}\right) \tag{3.7}$$

These definitions provide a convenient way to characterize performance, using only two parameters μ and σ. Given that

$$\Phi(Z) \equiv \int_{-\infty}^{Z} \phi(X)dX \tag{3.8}$$

and

$$\Phi\left(\frac{Z-\mu}{\sigma}\right) \equiv \int_{-\infty}^{\frac{Z-\mu}{\sigma}} \phi(X)dX = \left(\frac{1}{\sigma}\right)\int_{-\infty}^{Z} \phi\left(\frac{X-\mu}{\sigma}\right)dX \tag{3.9}$$

then

$$P_0(Z) = 1 - \Phi(Z) \tag{3.10}$$

$$P_1(Z) = 1 - \Phi(Z)\Phi\left(\frac{Z-\mu}{\sigma}\right) \tag{3.11}$$

These are the parametric representations for the false positive rate and true positive rate as P_0 and P_1, respectively. This model yields a direct calculation of localization accuracy using the two parameters.

$$P_{CL}(Z) = \left(\frac{1}{\sigma}\right)\int_{Z}^{\infty} \phi\left(\frac{X-\mu}{\sigma}\right)\Phi(X)dX \tag{3.12}$$

For N_0 non-target and N_1 target images, the estimation of all curves under Swensson's approach defines the likelihood function given by

$$\begin{aligned} \ln(L) = {} & N_0 \sum_{j=0}^{C-1} p_{0,j} \ln\left[\Phi(Z_{j+1}) - \Phi(Z_j)\right] \\ & + N_1 \sum_{j=0}^{C-1} \begin{pmatrix} cp_{1,j}\left(\frac{1}{\sigma}\right)\ln\left\{\int_{Z_j}^{Z_{j+1}} \phi\left(\frac{[X-\mu]}{\sigma}\right)\Phi(X)\,dX\right\} \\ +ip_{1,j}\ln\left\{\int_{Z_j}^{Z_{j+1}} \phi(X)\Phi\left(\frac{[X-\mu]}{\sigma}\right)dX\right\} \end{pmatrix} \end{aligned} \tag{3.13}$$

For each perceptual measurement bin (rating category) j, $cp_{1,j}$ and $ip_{1,j}$ are the fractions of correct and incorrect localizations over the total number of (target) choices, respectively, and $p_{0,j}$ is the decision rate in the non-target class. The method employs a least-squares fitting algorithm based on the Levenberg–

Marquardt approach, to simultaneously estimate the parameters μ and σ which maximize this likelihood function across the C rating categories. Once the two parameters are known, the ROC and LROC curves are completely specified.

Chapter 4

Mathematical Observers

Introduction

In the clinical setting, image analysis is always handled by a professional radiologist or photo interpreter. However, human observation may be impractical when it comes to large amounts of data to be processed, especially in the regimes of development and testing. Furthermore, they may not be able to use all the information as effectively as computers. Mathematical observers have been used frequently in lesion detection tasks; these include the ideal Bayesian logic observer, matched filter observers, and models based on linear discriminant analysis, referred to as the Hotelling observer. Further extensions of these models have included channelized filtering to model the frequency response of the human visual system.

There has been much interest in using mathematical observers to replace or influence human performance, (Chan, Doi et al. 1990; Doi, Giger et al. 1992) and much work has been conducted to develop numerical observers that accurately predict human results. There is much debate over which observer most accurately models human performance; different observers may be optimal in specific tasks.

Ideal Observer

We begin by considering the simplified binary classification task of an exactly known signal either present or not present on an exactly known background (SKE/BKE). As mentioned in the previous chapter, any observer performs a decision, regarding a given test image **g**, by extracting the relevant information and calculating some perceptual statistic. The observer then compares this statistic to a threshold value to determine which underlying class (signal present or absent) produced the test image.

The ideal observer is a term used to describe the best possible observer performance given the situation, through the optimal use of available information.[7] The ideal observer is not perfect; it only performs at the physical limit. It is subject to noise and uncertainty and is therefore expressed as statistical probability. It makes a classification decision based on the likelihood function, which is the probability ratio of the two classes. It compares this ratio, or test statistic, to a threshold to determine the outcome.

$$\lambda_{ideal}(\mathbf{g}) \equiv \mathbf{log}[L(\mathbf{g})] = \mathbf{log}\left[\frac{p(\mathbf{g}|\mathbf{g}_1)}{p(\mathbf{g}|\mathbf{g}_2)}\right] \tag{4.1}$$

The signal-to-noise ratio associated with lesion detectability is defined

$$d = \sqrt{\frac{\left[\boldsymbol{E}\left(\lambda(\mathbf{g}_2)\right) - \boldsymbol{E}\left(\lambda(\mathbf{g}_1)\right)\right]^2}{\frac{1}{2}\left[\mathbf{var}\left(\lambda(\mathbf{g}_1)\right) + \mathbf{var}\left(\lambda(\mathbf{g}_2)\right)\right]}} \tag{4.2}$$

where $\boldsymbol{E}\left(\lambda(\mathbf{g})\right)$ and $\mathbf{var}\left(\lambda(\mathbf{g})\right)$ are the conditional mean and variance of the test statistic for each of the two respective classes. The ideal observer however, is typically not a relevant model for lesion detection task, and furthermore, the log-likelihood may be difficult to determine. If there is randomness (noise) in the image, the log-likelihood can be highly non-linear in **g**, rendering the ideal observer mathematically intractable

---

[7] The ideal observer is often referred to as the Bayesian observer because the log-likelihood ratio follows Bayesian statistics.

(Barrett, Yao et al. 1993). Performance of the observer is typically rated as the detectability index ratio with that of the ideal observer.

Matched Filter

There are limitations on computer observers which exclude them from clinical tasks, as they must typically be trained using known samples from each class. For the binary task mentioned earlier, the observer model must have a sample of signal-present and signal-absent images. The mean difference between these classes yields the signal profile and is used as the template of the matched filter, or non-prewhitening (NPW), observer.

$$\lambda_{\mathrm{NPW}}(\mathbf{g}) = \mathbf{w}^t\mathbf{g} \tag{4.3}$$

where

$$\mathbf{w} = \bar{\mathbf{g}}_2 - \bar{\mathbf{g}}_1 \tag{4.4}$$

The test statistic for the matched filter is the inner scalar product of the signal profile and the image.

The SKE/BKE task is the simplest example of a signal detection task, and is important because the likelihood ratio can be calculated by simple linear discrimination. In this case, the only randomness will be within the observer measurement itself and can be modeled as a stationary, white, Gaussian process (Barrett, Gooley et al. 1992). In this situation, the matched filter is the ideal observer.

To illustrate this concept, is it helpful to constrain our problem to 2 dimensions. Figure 23 shows two distributions of the same Gaussian random process, in this case, the classifications of images into those with and without a signal present. Each scatter element represents an observer measurement made for a single image, i.e. the test statistic.

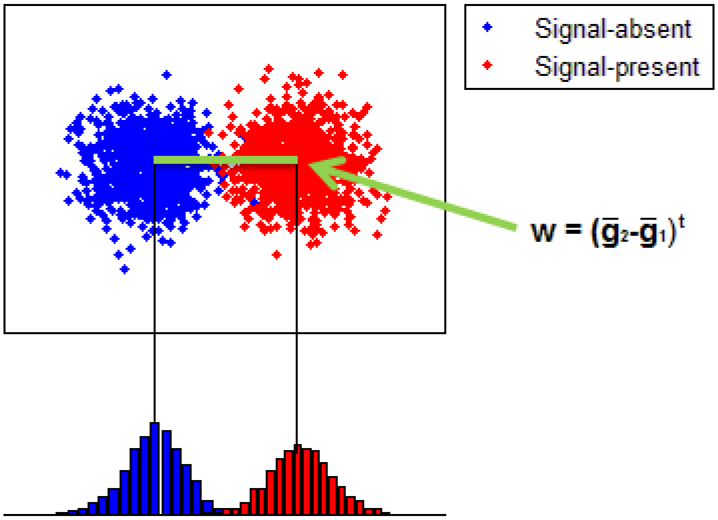


Figure 23. A random Gaussian process and the respective histogrammed density functions. The blue group represents the observer measurements made within samples from the signal-absent class and the red represents those from the signal-present class. The ideal observer template is defined by the difference in class means.

For the purpose here, it is practical to think of this as a set of two-pixel images, with a signal present or absent at the same pixel in each. The only difference in the distribution means lies along the dimension associated with that pixel. Thus the discriminant which provides maximum separation of the density functions is defined by the mean difference of that pixel, between the two classes. While this is an obvious oversimplification, it provides a convenient means to visualize the discriminative process of the observer model, and it is directly extended to situations in higher dimensions.

Correlated Noise

Raw PET data sets contain Poisson-structured noise, which is uncorrelated. When this raw data is reconstructed, using filtered backprojection or an iterative algorithm, it becomes correlated by the acquisition process, the use of frequency-selective filters, post reconstruction smoothing, and other image estimation methods. Now the measurement noise is Gaussian but not white, and the likelihood ratio is calculated through data whitening, in addition to the matched filter operation. This noise is modeled as an additive Gaussian process with the covariance matrices assumed to be identical between the two classes $j = 1, 2$.

$$\mathbf{K}_j \equiv \langle (\mathbf{g}_j - \bar{\mathbf{g}}_j)(\mathbf{g}_j - \bar{\mathbf{g}}_j)^t \rangle \tag{4.5}$$

The covariance matrix has the dimensions N by N, where N is the number of measurements. When **g** represents image data, N is the number of pixels.

In this case, the ideal observer first removes this correlation, and signal analysis is then performed. The pre-whitening operation is performed on the signal and the data by multiplying each with the square root of the inverse of the observer covariance matrix.

$$\lambda_{\mathrm{PW}}(\mathbf{g}) = (\bar{\mathbf{g}}_2 - \bar{\mathbf{g}}_1)^t \mathbf{K}^{-\frac{1}{2}} \mathbf{K}^{-\frac{1}{2}} \mathbf{g} = (\bar{\mathbf{g}}_2 - \bar{\mathbf{g}}_1)^t \mathbf{K}^{-1} \mathbf{g} \tag{4.6}$$

The SKE-BKE lesion detection task is unique in that the log-likelihood ratio of the ideal observer can (usually) be easily calculated. The assumption of identical class covariance matrices preserves the linearity in the calculation of the test statistic. This is illustrated below in Figure 24 and 25 following our earlier example. The distributions have the same means as before but now the measurements are correlated among pixels.

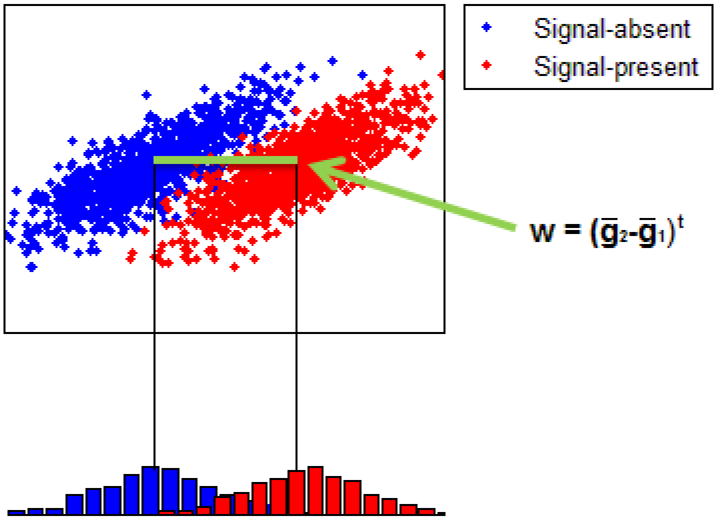


Figure 24. The discriminant defined by the difference in class means is no longer optimal when noise in the data is correlated.

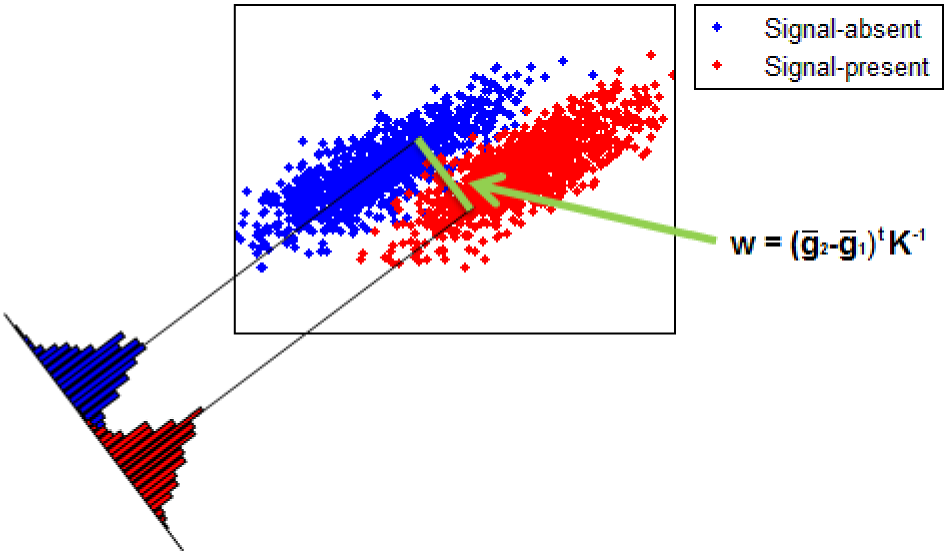


Figure 25. The distribution covariance is used with the mean difference to define the optimal discriminant function in correlated noise.

This is not the case for real data however, which has randomness not just in the measurements, but within the objects being imaged themselves; a full-rank class ensemble covariance accounts for these variations.

Several groups have shown that human efficiency in correlated noise is only about 20% relative to the ideal observer (Judy, Swensson et al. 1981; Burgess 1985). Further, Myers el al. evaluated human performance using SKE-BKE data with a post-detection filter to introduce different degrees of noise correlation (Myers, Barrett et al. 1985). This method generated a range of noise power spectra of the form ρ^n, where ρ is the spatial frequency and n = 1,2,3,4. This study found that human efficiency degrades rapidly as the degree of noise correlation increases, suggestive that humans are incapable of performing the prewhitening operation.

It should be noted that the square covariance matrix **K** must be full-rank in order to be invertible. This means that the number of random vectors in the sample must equal or exceed the number of elements (pixels) in each vector, i.e. there must be at least N samples in **g**. This requirement is nearly always impossible to achieve with images of any reasonable size.

Linear Discriminants

In the SKE-BKE task, the log likelihood is a first-order function in **g** and relatively easy to compute. However, when this test statistic becomes nonlinear, it becomes difficult or even impossible to compute. Through the use of linear discriminants, the test statistic can be approximated providing a work-around for the mathematical insolvability of nonlinear likelihood equations. The optimal choice of linear discriminant function maximizes performance of the classification task based on training data (Fisher 1936). The linear discriminant test statistic has the form

$$\lambda_{\mathrm{lin}}(\mathbf{g}) = \mathbf{u}^t\mathbf{g} \tag{4.7}$$

In 1931 Harold Hotelling introduced his T^2 value as a measure of class separability, which estimates the amount of the difference in the means of the classes (Hotelling 1931). The optimal linear discriminant will maximize the T^2 value, and so these models are termed Hotelling observers.

Calculations of Fisher's discriminant and Hotelling's T^2 assume complete knowledge of the ensemble distribution properties. To overcome this, interclass ($\mathbf{S}_1$) and intraclass ($\mathbf{S}_2$) scatter matrices are calculated (shown below for the binary classification problem)

$$\mathbf{S}_1 = \boldsymbol{P}_1 \boldsymbol{P}_2 (\bar{\mathbf{g}}_2 - \bar{\mathbf{g}}_1)(\bar{\mathbf{g}}_2 - \bar{\mathbf{g}}_1)^t \tag{4.8}$$

and

$$\mathbf{S}_2 \equiv \sum_{k=1}^{2} \boldsymbol{P}_k \mathbf{K}_k = \sum_{k=1}^{2} \boldsymbol{P}_k \langle (\mathbf{g} - \bar{\mathbf{g}}_k)(\mathbf{g} - \bar{\mathbf{g}}_k)^t \rangle_k \tag{4.9}$$

The interclass scatter matrix $\mathbf{S}_1$ expresses the average distance between the distribution means of the two classes, and has rank 1 for our binary task (k-1). The intraclass scatter matrix $\mathbf{S}_2$ expresses the average class covariance matrix within the data, found by summing the probability-weighted ($\boldsymbol{P}$) covariance matrices of both classes. Its rank is determined by the number of samples used in the calculation.

In terms of these scatter matrices, a measure of class separability called the Hotelling trace J is defined as (Smith and Barrett 1986)

$$J = \mathrm{tr}(\mathbf{S}_2^{-1}\mathbf{S}_1) \tag{4.10}$$

The matrix $\mathbf{S}_2^{-1}\mathbf{S}_1$ has the same rank as $\mathbf{S}_1$ and the Hotelling trace sums its diagonal elements.[8] The construction of J is intuitive; as it increases, so does the relative separation between the class means, while at the same time, the spread within each class is minimized. This scalar measure has the same form as Hotelling's T^2 value except that sample statistics are used in the place of ensemble ones.

This is closely related to the Fisher criterion, where we have

$$\mu(\mathbf{u}) = \frac{\mathbf{u}^{\mathrm{T}}\mathbf{S}_1\mathbf{u}}{\mathbf{u}^{\mathrm{T}}\mathbf{S}_2\mathbf{u}} \tag{4.11}$$

---

[8] The diagonally elements of this matrix are the eigenvalues of the discriminant functions. Since the rank of $\mathbf{S}_2^{-1}\mathbf{S}_1$ is (k-1), there will be (k-1) eigenvectors and non-zero eigenvalues.

where $\mathbf{u}$ is the optimal discriminant function. We find $\mathbf{u}$ by maximizing μ (setting $\frac{d}{d\mathbf{u}}[\mu(\mathbf{u})] = \mathbf{0}$) which gives us

$$\mu\mathbf{u} = \mathbf{S}_2^{-1}\mathbf{S}_1\mathbf{u} \tag{4.12}$$

The optimal linear discriminant for the binary classification problem, found by solving this general eigenvalue problem, is

$$\mathbf{u} = (\bar{\mathbf{g}}_2 - \bar{\mathbf{g}}_1)^t\mathbf{S}_2^{-1} \tag{4.13}$$

and the test statistic associated with J is

$$\lambda_{\text{Hot}}(\mathbf{g}) = (\bar{\mathbf{g}}_2 - \bar{\mathbf{g}}_1)^t\mathbf{S}_2^{-1}\mathbf{g} \tag{4.14}$$

This exactly equals the ideal prewhitening matched filter (PWMF) observer in SKE-BKE tasks, when the observer covariance matrices are equal for both classes. In the case of normal class distributions, the Hotelling observer is equal to the ideal observer.

An attempt to directly calculate the full ensemble statistics from sample training data fails in practice however since the number of sample images is typically insufficient to form a full-rank $\mathbf{S}_2$ matrix. Solutions to this singularity problem have been proposed that involve simulating noise-free images or using iterative algorithms to estimate the discriminant function (Fiete, Barrett et al. 1987). It is also possible to reduce the matrix dimensionality by extracting frequency-specific features of the data.

Human Response

As previously mentioned, Myers et al. used SKE-BKE data to test whether human observers are able to perform the prewhitening operation (Myers, Barrett et al. 1985). Human observer detectability fell off with increased correlated image noise, while that of the ideal (prewhitening) observer remained constant. This suggested the use of the non-prewhitening matched filter (NPWMF) to model human performance. The human observer has also been modeled as a matched filter without the prewhitening operation, but the two have shown poor agreement in various tasks (Yao and Barrett 1992). For example, in the case of a known signal in a non-uniform background, the NPW observer performance falls off with background lumpiness while human performance does not follow this trend (Rolland and Barrett 1992).

The inconsistencies have been explained by internal noise present within the human observer (Burgess and Colborne 1988). While human performance could be predicted by the NPWMF for some tasks in both correlated and uncorrelated noise, it could not always be accurately predicted by a generic observer. A modified model was needed to predict human performance in all tasks, one that incorporated physiological effects of the human detection system.

The incorporation of a single spatial-frequency eye filter and an internal noise mechanism in the Hotelling observer showed closer agreement was with the human observer (Loo, Doi et al. 1984; Giger and Doi 1987; Barrett, Yao et al. 1993). Further, it is understood that the human visual system is composed of multiple narrow spatial-frequency filters. This psychophysical model is also incorporated in the NPWMF to more accurately predict human observer performance.

Response Channels

Contrast sensitivity of an observing system is defined as the minimum contrast ratio that is able to be detected. Visual performance, rated by its contrast sensitivity function, measures the overall sensitivity to sinusoidal pattern contrast as a function of spatial frequency (i.e. cycles/visual degree). Much work has been done investigating the performance of the human visual system (Campbell and Robson 1968). It is well known that the ability of humans to detect patterns is based on a global response to a pattern, rather than the local response of a single optical nerve. However, certain cells are uniquely tuned to identify specific patterns and frequencies.

This physiological effect is modeled by computers through channelized filters. A channel is an independent processor which acts as a band-pass filter tuned to a finite frequency bandwidth. The component with the highest detectability is then identified when it rises above threshold, instead of analyzing the data across a large continuum of frequencies, which masks weaker effects and raises the signal detectability threshold. A robust model observer capable of modeling human performance in all tasks was achieved through the incorporation of channels into the observer model (Myers and Barrett 1987; Barrett, Yao et al. 1993; Burgess, Shaw et al. 1999; Gifford, King et al. 2000). A channel, as it applies to mathematical observers, is represented by a vector that describes the channel impulse response centered on the tumor location (Gifford, King et al. 2000). They are defined in frequency space (pixels^{-1}) and the responses are their inverse Fourier transforms.

Channels are chosen for different purposes. If the goal is to optimize the hardware of the acquisition system and the calculation of the likelihood of each class is intractable, the ideal linear observer model would be suitable. However if the goal is to assess the performance of different reconstruction algorithms, the ideal observer model may not useful because it is invariant to invertible image processing (Barrett and Myers

2004). In any case, channels are implemented to handicap the observer, and in so doing, the dimensionality of the data is reduced.

The 2-class channelized Hotelling observer (CHO) intraclass scatter matrix takes the form

$$\mathbf{S_2}_{\mathrm{ch}} = \frac{1}{2}\mathbf{U}^{\mathrm{t}}[\mathbf{K_1} + \mathbf{K_2}]\mathbf{U} \tag{4.15}$$

The data covariance matrix for a 2-dimensional image with N pixels has the dimensions N × N, yielding a matrix containing a number of elements on the order of 10^4-10^5 for a typical size image; this size would be bigger if dealing with 3-dimensional volume arrays. Processing the images with a channelized filter reduces the dimensionality of the problem, and the M-channel covariance matrix now has the dimensions of M × M, which is significantly smaller. The CHO prewhitening operation decorrelates the image noise between the channel responses, instead of the pixels.

The matched filter of the CHO uses the signal as it is processed through the channels to give us the complete template

$$\mathbf{w}_{\mathrm{cho}} = (\bar{\mathbf{g}}_2 - \bar{\mathbf{g}}_1)^t \mathbf{U}\, \mathbf{S_2}_{\mathrm{ch}}^{-1} \mathbf{U}^t \tag{4.16}$$

The SNR of the CHO is the square root of the inner product of the template and the mean signal, namely

$$\mathrm{SNR}_{\mathrm{cho}} = \left[(\bar{\mathbf{g}}_2 - \bar{\mathbf{g}}_1)^t \mathbf{U}\, \mathbf{S_2}_{\mathrm{ch}}^{-1} \mathbf{U}^t (\bar{\mathbf{g}}_2 - \bar{\mathbf{g}}_1)\right]^{\frac{1}{2}} \tag{4.17}$$

The SNR-rated performance of a mathematical observer is compared with that of a human observer using the relationship between the AUC and the detectability index.

Chapter 5

Lesion Detection

Introduction

The idea to use time-of-flight (TOF) information from coincident photon detections to improve PET image quality has been explored before (Tomitani and Tanaka 1980; Budinger 1983), and it has recently been revived due to advances in PET scanner technology (Conti, Bendriem et al. 2005). Additionally, the spatially-variant point spread response (PSF) of a PET system can be incorporated into the system model (Panin, Kehren et al. 2006; Tong, Alessio et al. 2009; Walker, Julyan et al. 2009; Daube-Witherspoon, Matej et al. 2011).

The clinical benefit of TOF and PSF in PET reconstructions is of major interest. In oncology, lesion detectability is important in the early diagnosis and staging of patients. Positive patient prognoses are strongly dependent on early detection, so the ability to detect small, low-intensity lesions is critically important.

Incorporation of these reconstruction models has great potential to improve image quality in oncological studies (Kadrmas, Casey et al. 2009; Armstrong, Williams et al. 2011). Recent work has reported improvement from TOF due to the faster iterative convergence (Karp, Surti et al. 2008; Lois, Jakoby et al. 2010), and from PSF due to more uniform spatial resolution which enhances small foci of uptake, especially at large radial distances (Casey 2007; Tong, Alessio et al. 2010). Little work has been done, however, to *objectively* assess the impact of these benefits for lesion detection in clinical whole-body ^{18}F-FDG oncological studies (Surti, Scheuermann et al. 2011). This is partly because of the complications in translating realistic tasks into the experimental setting and the difficulty in performing a study on sufficient patient numbers to ensure statistical power (El Fakhri, Surti et al. 2011).

A rigorous evaluation methodology was developed to quantify the benefit of these new reconstruction models in a real lesion detection task using both human and mathematical observers (Schaefferkoetter, Casey et al. 2013). Three physicians (2 board-certified nuclear medicine physicians and 1 radiologist certified in nuclear medicine) were recruited for the study. The use of actual patient data accounted for the wide range of realistic subject volumes and physiological variability that exist in the clinical setting; the addition of known signals provided a gold standard for the detection task.

Additionally, the performances of eight mathematical observer models will be evaluated in a localization task and compared to those of the physicians. Non-prewhitening and Hotelling (with internal noise) schemes are used, each with 4 channel configurations.

Background

This work aims to assess the added utility of both TOF and PSF reconstructions in the detection of lesions of varying contrast placed at different anatomical locations. The hypothesis is that the enhancement of the visibility of small foci when employing the combination of both models is superior to that when using either independently or neither (Kadrmas, Casey et al. 2009), and that these effects are amplified for weaker signals and in larger patients (Surti and Karp 2009).

Phantoms are used to measure detectability performance because the physical geometry and exact concentration of the activity distribution is known. However, this neglects many of the complexities that exist in the clinic. The current experiment implements a recent method of statistically inserting small spherical volumes of known activity into actual patient scans (El Fakhri, Santos et al. 2007), providing a rigorous, highly reproducible method of testing.

The combination of real clinical data with that of simulated lesions reduces the complexities associated with a problematical clinical evaluation to those of a simple signal detection task, in which the signal is known statistically (SKS), but its location is not. It allows complete control over the signal to background contrast ratios, and it provides a gold standard for the classification task. This combines the control of a phantom and the realism of a patient to investigate the performance of different reconstruction algorithms.

A full range of contrast ratios will be analyzed using both human and mathematical observers, providing a wide range of observer confidence sampling for characterizing ROC performance. Furthermore the performance of the model observers will be compared to and validated by that of the physicians. This work is among the first studies to investigate the benefit of TOF in lesion-detection tasks using actual clinical data. This study is strongly motivated by the lack of research assessing the impact that the PSF model and TOF information has on improved patient care.

Previous Work

The current method of lesion simulation in patient data was developed by the group led by El Fakhri at Massachusetts General Hospital (MGH). It was validated through the use of a phantom study in which a 1.3cm-diameter sphere was filled with FDG and an elliptical phantom was also filled with a solution of FDG, with a contrast ratio of 2:1. The phantom was scanned both without and with the sphere positioned inside of it at an off-centered position. Then the sphere was scanned in air at the same position mounted to the grid then added to the phantom in sinogram space. The linear profiles of the images, as well as mean pixel noise and sphere-to-background contrast ratios were identical (El Fakhri, Santos et al. 2007).

This method was designed to study the image quality obtained from 2D versus 3D PET acquisition, in the context of lesion detection. It was found that the method of optimal acquisition is dependent on patient BMI, which was used as a surrogate for total patient attenuation. This study was conducted on a PET/CT scanner with BGO detectors. It concluded that 3D acquisition was superior for smaller patients until the individual's BMI reached 31; at this point the increased sensitivity of 3D WB-PET to the scattered and random coincidences led to noisier images and lesion detectability deteriorated. These conclusions were based on CHO-SNR. This lesion detection study will build on the methods developed by El Fakhri et al. to assess image quality in a clinically relevant paradigm.

Several groups have recently investigated the benefit of TOF information on image quality. One work (Karp, Surti et al. 2008) used simulations and phantom data, with TOF information alongside iterative reconstruction, to demonstrate that the benefit of TOF could not be measured by a single gain factor. The addition of the TOF information in the EM algorithm increases the image convergence rate, leading the reconstruction to reach a higher contrast ratio in fewer iterations. Furthermore they found that the TOF reconstruction converges more uniformly and to different contrast values than conventional reconstructions.

Lois et al. (Lois, Jakoby et al. 2010) compared PSF reconstructions, with and without TOF. An image quality phantom and the clinical data from a small subset of patient scans were used in the subjective assessment of image quality by a physicist and a nuclear medicine physician to establish optimal reconstruction parameters. The data from a cohort of 100 patient scans[9] were then reconstructed according to these parameters and SNR statistics were calculated at various lesions located at different anatomical sites. This group found that TOF information leads to faster convergence of the reconstruction algorithm, better image contrast, and lower image noise. As expected, the SNR gain due to TOF had the greatest effect in patients with higher BMI compared to non-TOF, because of its larger area of uncertainty in event localization and propagated noise.

The impact of TOF on lesion detectability has also been explored to an extent. One such study compared TOF improvement in SKE/BKE lesion detection using a uniform phantom and spherical lesions of varying sizes (Surti and Karp 2009). A NPWMF observer was used to quantify the SNR gain in performance. It compared performance, based on a fixed scan time and iteration number. This group reported faster

---

[9] These are the patient data sets used in the current study.

convergence and thus higher contrast recovery with TOF, leading to improved lesion detection, especially for shorter scan times. This group noted that TOF-PET will have maximum benefit in the detection of small lesions in large patients.

Another paper from the same year compared four reconstructions, LOR-OSEM with PSF, TOF, and a PSF+TOF (Kadrmas, Casey et al. 2009). Synthetic lesions were physically added to an anthropomorphic phantom, and mathematical as well as human observers were used to gauge detection performance. It used LROC analysis, and the numerical observer was a channelized non-prewhitening (CNPW) observer with a 10-channel, dense difference-of-Gaussian (D-DOG), configuration. This group noted that TOF- PET will have maximum benefit in the detection of small lesions in large patients. They described the benefit gained from TOF as similar to that of the PSF correction, seen in the reconstructed images as the sharpening of small focal lesions, especially at locations far from the center. However the cumulative benefit of both PSF *and* TOF significantly outperformed either if used alone in lesion detection tasks. We expect similar results in the current work.

The group at MGH led by El Fakhri conducted a recent experiment to evaluate improvement from TOF in lesion detection across many different parameters, including scan time, lesion contrast, body mass index (BMI), and anatomical location (El Fakhri, Surti et al. 2011). This group employed the current method of adding simulated lesions to actual patient data and CHO-SNR was used to rank performance. This team reported greatest gains in performance using TOF for shorter scan times, smaller lesions, and larger patients. These results were expected since reconstruction parameters were fixed for both reconstructions. This study is similar to the current one in that actual patient data were used, and performance is compared between variations in patient, lesion, and scanning parameters. The key differences of the current study include the optimization of the individual reconstructions and the use of board certified radiologists to validate the methodology. Also, performance of the PSF-model was not considered in this previous work.

Materials and Methods

Patient Population

Ninety-six patients underwent routine clinical PET/CT scans at the University of Tennessee Medical Center after 90-min uptake of 370 MBq of ^{18}F-FDG according to standard clinical protocol (Lois, Jakoby et al. 2010).

The BMI as defined by the National Institutes of Health publications (Tarantola, Zito et al. 2003) was used to classify the patient population. In our population, BMI ranged from 16.14 to 55.61 (mean 28.51). Of these patients, 40 were selected for analyses by the numerical observers, including the population of 33 disease-free bed positions from 23 patients (BMI 20.2-46.7, mean 30.9, std. dev. 6.34) selected for the human observer study.

PET Acquisition

All patients were scanned on a Biograph PET/CT, a TOF-capable, fully 3D PET scanner together with a 6-slice Sensation spiral CT (Siemens Molecular Imaging). The PET system comprised 4 rings of lutetium oxyorthosilicate (LSO) crystal detectors. The 60 cm transverse FOV was defined by a coincident acceptance window of 4.1 ns, with an energy threshold of 435-650 keV. The acceptance window was divided into 312 ps timing bins for the TOF measurements. The energy and timing resolution were measured with an ^{18}F line source resulting in 11.7% (FWHM) and <550 ps (FWHM), respectively. The average spatial resolution (FWHM), measured at 1 cm and 10 cm from the center of the transverse FOV, was 4.4 mm and 5.2 mm transaxially, and 4.4 mm and 5.8 mm axially (Jakoby, Bercier et al. 2008).

Patients underwent a whole-body (WB) PET/CT consisting of 4 to 7 bed positions, each scanned for 2 to 4 min, and PET data were acquired in list-mode and reconstructed. The volumes were reviewed, and locations of possible lesion sites were recorded. Forty locations were chosen including different anatomical regions of interest (e.g., lungs, soft tissues, bones). This process intended to simulate focal uptake consistent with pathology, such as solitary nodules or small metastatic deposits.

To model the artificial lesions, a 1-cm inner-diameter ^{18}F-filled sphere was separately scanned in air at each of the 40 predetermined locations. This was accomplished by mounting the sphere on a grid, shown in Figure 26, with over 600 threaded holes and adjusting the horizontal bed position to match the z-coordinate of the location.

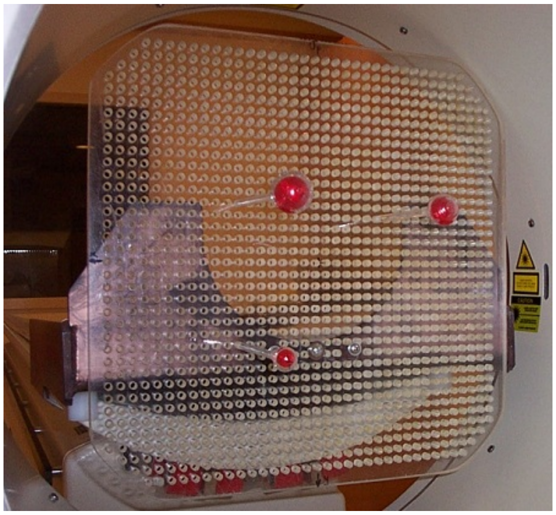


Figure 26. Grid used to mount artificial "lesions" in the PET field of view

This allowed exact placement of the lesion at any desired location in the PET scanner. The positions of the spheres were noted for use by the numerical observer. Both patient and sphere data were acquired in 3D mode. Randoms were estimated using the delayed-window subtraction method, and smoothed prior to subtraction from the total prompts. We assumed that the scatter associated with the sphere acquired in air was negligible compared to the scatter associated with the patient. The 40 locations were scanned separately for 3 minutes each, and list-mode data stored. Lesions could have been added to the patient data using simulations, but to ensure realistic situations, we chose to acquire the "lesion" data on the scanner to avoid errors associated with normalization and incomplete modeling of the scanner components.

Lesion Addition

Acquisition data were collected in list-mode so both conventional and TOF sinograms could be formed. The patient and sphere data were binned into 3D sinograms, with dimensions: 336 radial bins, 336 azimuthal bins and 559 3D planes (109 direct planes and 450 oblique planes grouped in 6 segments). The prompt and random events were contained within separate sinograms. The TOF data from the tomograph were subdivided into 15 time bins, each 312 ps wide; the conventional sinograms ignored the TOF measurements, combining all time bins into one.

For each lesion-present image, a “lesion” sinogram from one sphere (neglecting randoms) was forward attenuated by dividing it by the attenuation map calculated from the patient CT data. The attenuated lesion sinogram was then added to that of the patient. Figure 27 shows this basic scheme.

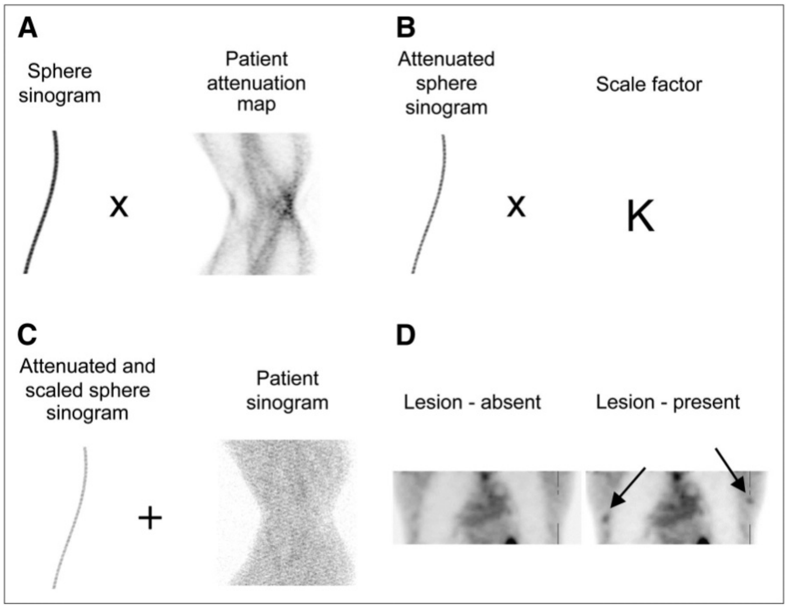


Figure 27. The process of simulating realistic lesions in patient scans.

To highlight the performance differences across the reconstructions, the lesion data were scaled to produce contrast values that would yield just marginal detectability in the baseline images. This way they would be enhanced with TOF and PSF, leading to different observer detectability results. Previous work has reported higher detectability of liver versus lung lesions, at the same contrast, due to the increased noise in the low-count regions (El Fakhri, Surti et al. 2011). We chose to scale the lesions in the lung to higher relative contrasts to account for the increased noise.

Lesion contrast, defined as

$$Contrast = \frac{(Signal - Background)}{Background} \tag{5.1}$$

determined the activity level of the simulated lesions. For each patient-lesion combination, a spherical region of interest (ROI) was drawn on the lesion-free image at the location where the lesion was to be inserted. The lesion was then scaled to an appropriate activity level defined by the mean value of this background ROI. This was done manually for several patients until a correlation was found between background ROI statistics and desired lesion contrast.

Table 4. Distributions of mean lesion contrasts, grouped by reconstruction method

Reconstruction	***Liver***	***Lungs***
OP	1.27	2.01
PSF	1.36	2.46
TOF	1.32	2.24
TOF+PSF	1.46	2.99

To generate a large population of test images, the lesion strength and scaling factor correlation, found in the limited subset, was applied to all patient-lesion combinations. Each of the lesion-present final images was reconstructed twice, the first pass used the scaling calculated from the background site and the second accounted for scaling adjustments if needed. The distribution of lesion contrasts ranged from 1.09 to 9.21 (mean 1.84, std.dev 0.91) across the four reconstructions. Figure 28 shows the distribution, pooled and separated by reconstruction method

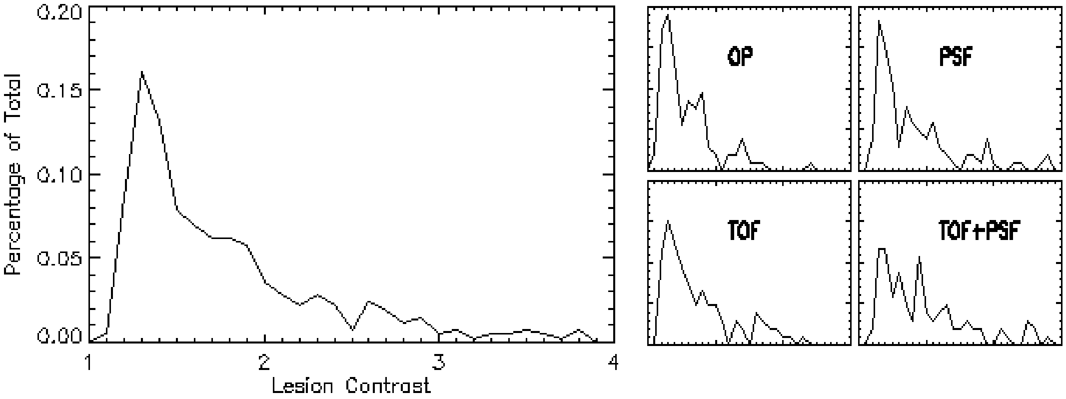


Figure 28. Distributions of lesion contrasts used in observer study

Mathematical Observer Models

The channelized Hotelling observer (CHO), with sparse difference-of-Gaussian (S-DOG) channel configuration (three radially symmetric channel profiles), has shown good agreement with human performance (Yao and Barrett 1992; Barrett, Yao et al. 1993; Abbey and Barrett 2001). A channelized non pre-whitening scheme with a dense difference-of-Gaussian (D-DOG) channel configuration (ten channel profiles) has also been used (Gifford, Kinahan et al. 2007; Kadrmas, Casey et al. 2009; Kadrmas, Casey et al. 2009); this work includes both models.

The DOG channels are defined by

$$C_n(\rho) = e^{-\frac{1}{2}\left(\frac{\rho}{Q\sigma_n}\right)^2} - e^{-\frac{1}{2}\left(\frac{\rho}{\sigma_n}\right)^2} \tag{5.2}$$

The set of σ_n is defined according to

$$\sigma_n = \sigma_0 \alpha^n \tag{5.3}$$

As in previous works (Gifford, King et al. 2000; Abbey and Barrett 2001; Gifford, Kinahan et al. 2007), the values of Q and α are both set to 2. The σ_n sets the standard deviation of the two Gaussian curves in frequency domain. Setting σ_0 to 0.015 (Abbey and Barrett 2001), the channel profiles are seen in Figure 29.

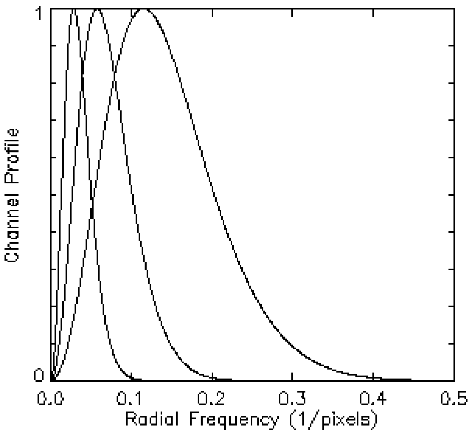


Figure 29. Channel profiles for the S-DOG configuration

The corresponding spatial channels are obtained through an inverse Fourier transform. The linearity of the transform allows us to handle each part in Equation (5.2) separately. For a general Gaussian function, the inverse Fourier transform is

$$\mathcal{F}_\rho^{-1}\left[e^{-\frac{1}{2}\left(\frac{\rho}{\sigma_n}\right)^2}\right](r) = \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{\rho}{\sigma_n}\right)^2} e^{2\pi i \rho r}\, d\rho \tag{5.4}$$

Following Euler's relationship, the complex exponential can be expanded.

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{\rho}{\sigma_n}\right)^2} e^{2\pi i\rho r}\, d\rho$$

$$= \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{\rho}{\sigma_n}\right)^2} \cos(2\pi\rho r) d\rho + i \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{\rho}{\sigma_n}\right)^2} (2\pi\rho r) d\rho \qquad (5.5)$$

The second integrand contains the product of the sin function and a zero-mean Gaussian. We can therefore eliminate the complex portion of the formula, since integrating an odd function over a symmetric range yields zero. The value of the first integral is found through the identity (Abramowitz and Stegun 1972)

$$\int_{0}^{\infty} e^{-\alpha t^2} \cos(2xt) dt = \frac{1}{2}\sqrt{\frac{\pi}{\alpha}} e^{-\frac{x^2}{\alpha}} \qquad (5.6)$$

So,

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{\rho}{\sigma_n}\right)^2} \cos(2\pi\rho r) d\rho = \sqrt{2\pi}\sigma_n e^{-2(\pi\sigma_n r)^2} \qquad (5.7)$$

Thus, the spatial channels as defined by Equation (5.2) are

$$C_n(r) = Q\sigma_n e^{-2(\pi Q\sigma_n r)^2} - \sigma_n e^{-2(\pi\sigma_n r)^2} \qquad (5.8)$$

The factor of $\sqrt{2\pi}$ has been omitted since the relative performance of the mathematical observer is invariant to scale transformations. The Fourier transform is represented in Figure 30 for 1 dimension, and the spatial profiles are then shown in 2D in Figure 31.

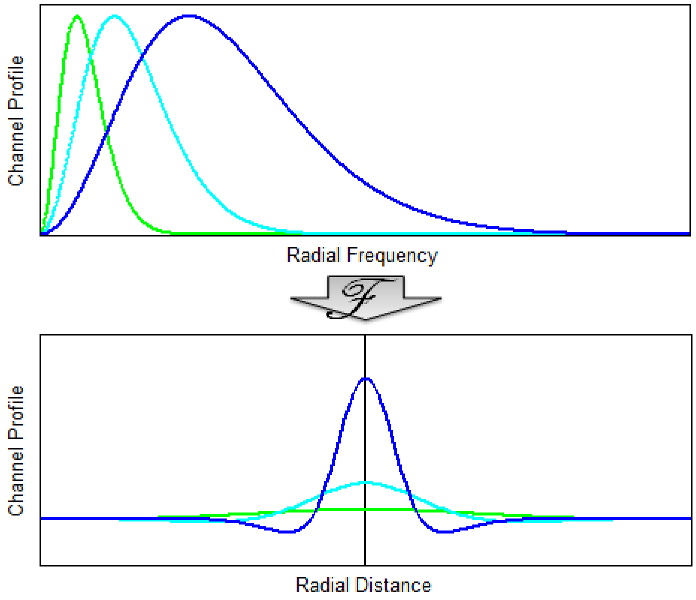


Figure 30. S-DOG channel profiles, transformed from frequency to spatial domain

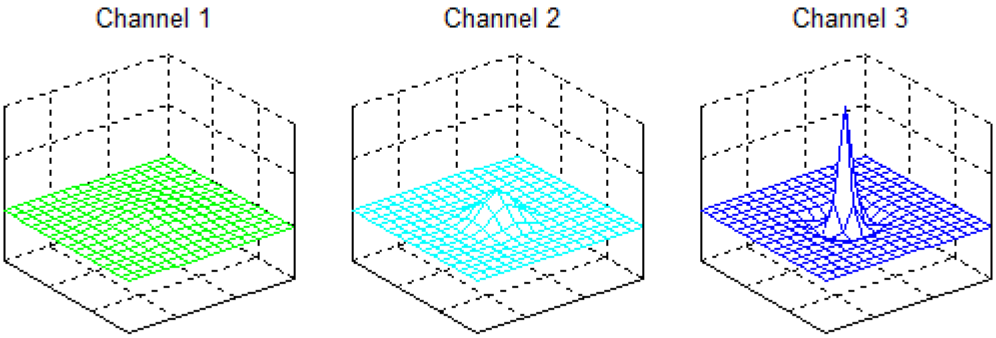


Figure 31. S-DOG spatial channel profiles shown in 2 dimensions

The projection of the (sub) image onto the set of spatial channels given by Equation (5.8) produces a scalar response, effectively reducing the dimensionality of the problem. We can model the discrimination task of the three-channel observer using spatial coordinates. Processing a sample set of images through the channelized model yields a three dimensional vector for each. The sample images are plotted below in Figure 32.

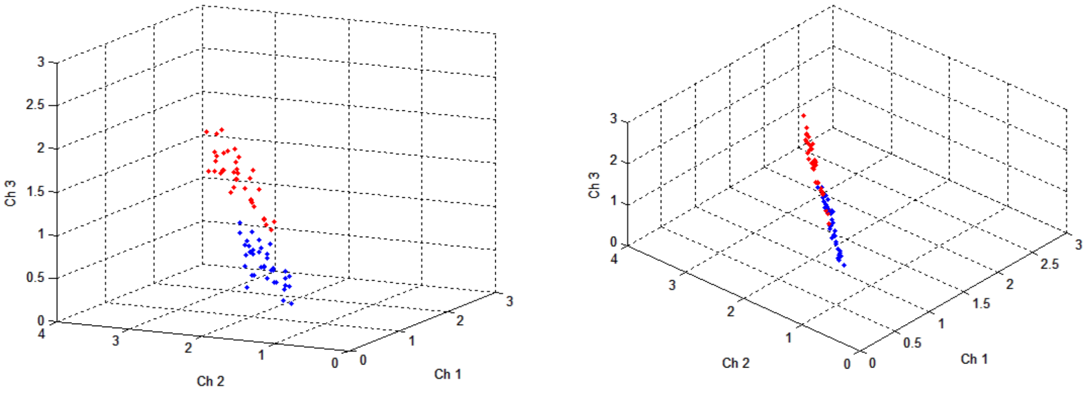


Figure 32. The sample images represented as filtered data. The responses to the channels are correlated.

Figure 32 shows 76 images taken from the liver, represented by their (scaled) three-dimensional channelized responses. The sample consists of 38 lesion-present images (red) and 38 lesion-absent images (blue). A view from another angle shows the correlation introduced by the filtering, especially obvious between the lower frequency channels.

The channelized non pre-whitening (CNPW) observer uses a discriminant function defined by the difference in class means.

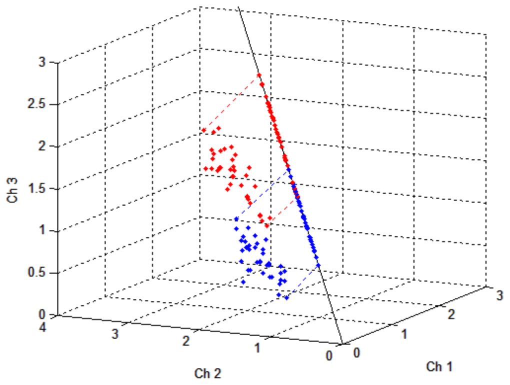


Figure 33. In the CNPW model, the class means define the discriminant function.

In the CHO model, the intraclass scatter matrices are used with the mean difference to calculate the linear discriminant. This function is the ideal projection for the binary classification task, under the assumption that each class shares the same normal distribution. This scheme is similar to Fisher linear discriminant analysis, except that the full ensemble statistics are not known, but rather approximated from the sample distributions.

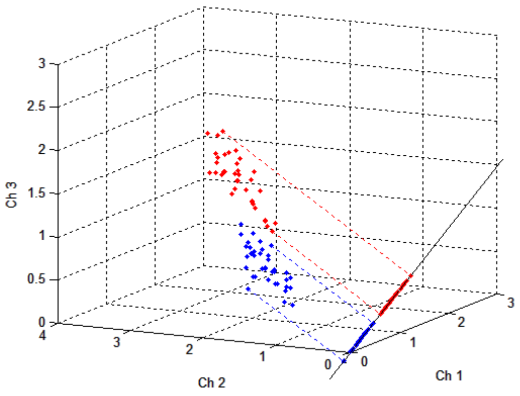


Figure 34. In the CHO model, the class means and intraclass scatter matrices define the discriminant function

The CHO maximizes the separation of classes, while minimizing the spread within each class. The performance of this ideal model often disagrees with that of a human, due in part to internal noise within the human observer. This noise is presumed to be similar to the variance within each frequency channel, and its effect on the classification task is accounted for by doubling the magnitudes of the diagonal elements of the (channelized) prewhitening matrix (Abbey and Barrett 2001; Gifford, King et al. 2005). In the current example, the new discriminant function looks like

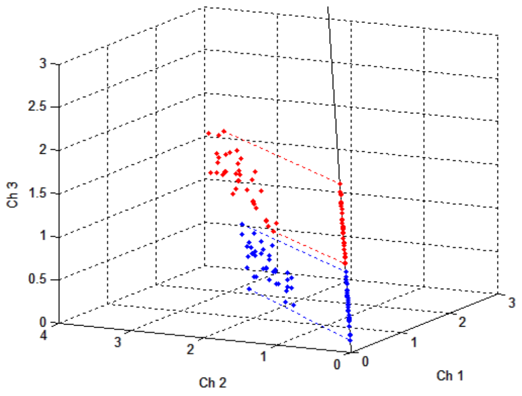


Figure 35. The new CHO discriminant function resulting from observer noise added to the intraclass scatter matrix.

This “handicapped ideal” discriminant function necessarily yields a lower value of J (class separation) and has shown better agreement with human performance. It is noted that, after the covariance matrix has been adjusted, Equation (4.17) is no longer valid and SNR must be calculated according to Equation (4.2)

$$SNR = \sqrt{\frac{(\bar{\mu}_1 + \bar{\mu}_2)^2}{\frac{1}{2}(\sigma_1^2 + \sigma_2^2)}} \tag{5.9}$$

where μ is the observer test statistic for each image and σ^2 is the variance of the test statistics in each distribution (target and non-target). In this study, the CHO is implemented with internal noise to more closely match human performance.

Imaging Conditions

The mathematical observer SNR measured the (binary) class separability and was used to quantify reconstruction performance for various lesion contrasts (in the liver), patient scan times, and patient BMI. Lesion contrast was varied by adjusting the signal activity within each patient; 20K, 40K, and 60K counts were extracted from the sphere list-mode data and added to the data from 25 patients with 3-minute clinical scans, yielding mean contrast values of 1.14 ± 0.03, 1.30 ± 0.07, 1.47 ± 0.11, respectively. The effect of TOF and PSF on patient scan time has been investigated before (Kadrmas, Oktay et al. 2012) and similar analyses were performed here; one , two, and three-minute scan time realizations were extracted from the patient data sets. To evaluate the influence of patient size on performance, the original 40 patients were grouped by BMI. Patients with BMI below 27 and above 29.5 were pooled into low and high BMI classes, respectively, producing two separate groups of 18 patients. For all conditions, the mean gains in observer SNR were evaluated over 10 iterations of the reconstruction algorithms.

Reconstructions

This work considered iterative reconstruction methods, as these are the current standard for clinical imaging. The ordinary Poisson (OP) OSEM (Shepp and Vardi 1982; Michel, Sibomana et al. 1998) was the baseline reconstruction. This reconstruction treats the data as a discrete number of Poisson variables and iterates towards the maximum value of the likelihood function. In this scheme, the Poisson statistics are preserved.[10]

For the images used in the observer study, the image processing schemes were optimized on the basis of performance of a model observer in detecting the presence of a 1-cm spherical signal in a random, noisy background. Single-bed PET data were taken from 40 patients, both with and without a lesion added in the liver, and reconstructed

---

[10] Some algorithms interpolate the projections and cast them into equally spaced radial bins. This compromises the Poisson nature of the data and can lead to less optimal reconstructions, especially in low count-rate conditions.

with each of the 4 algorithms (OP, PSF,TOF, and TOF+PSF), from 1-10 iterations, using 12 subsets, with 14 azimuthal angles per subset. According to clinical protocol of the institution, the sinograms were rebinned radially and angularly to 168 x 168; the final image matrix was also 168 x 168, with a pixel size of 4.073 mm, and a slice thickness of 2.027 mm. These images were then post-processed both without smoothing and with Gaussian kernels ranging in size from 4-10 mm (FWHM), in 1 mm increments, for a total of 25,600 images.

The numerical observer was the 2D D-DOG CNPW (Gifford, Kinahan et al. 2007) in which a 15 x 15 pixel sub-image, centered on the lesion site, was processed through ten frequency-selective channels. The maximum observer SNR of the observer (Equation 5.4) was used to determine both the optimal iteration number and smoothing kernel size to be used for each reconstruction.

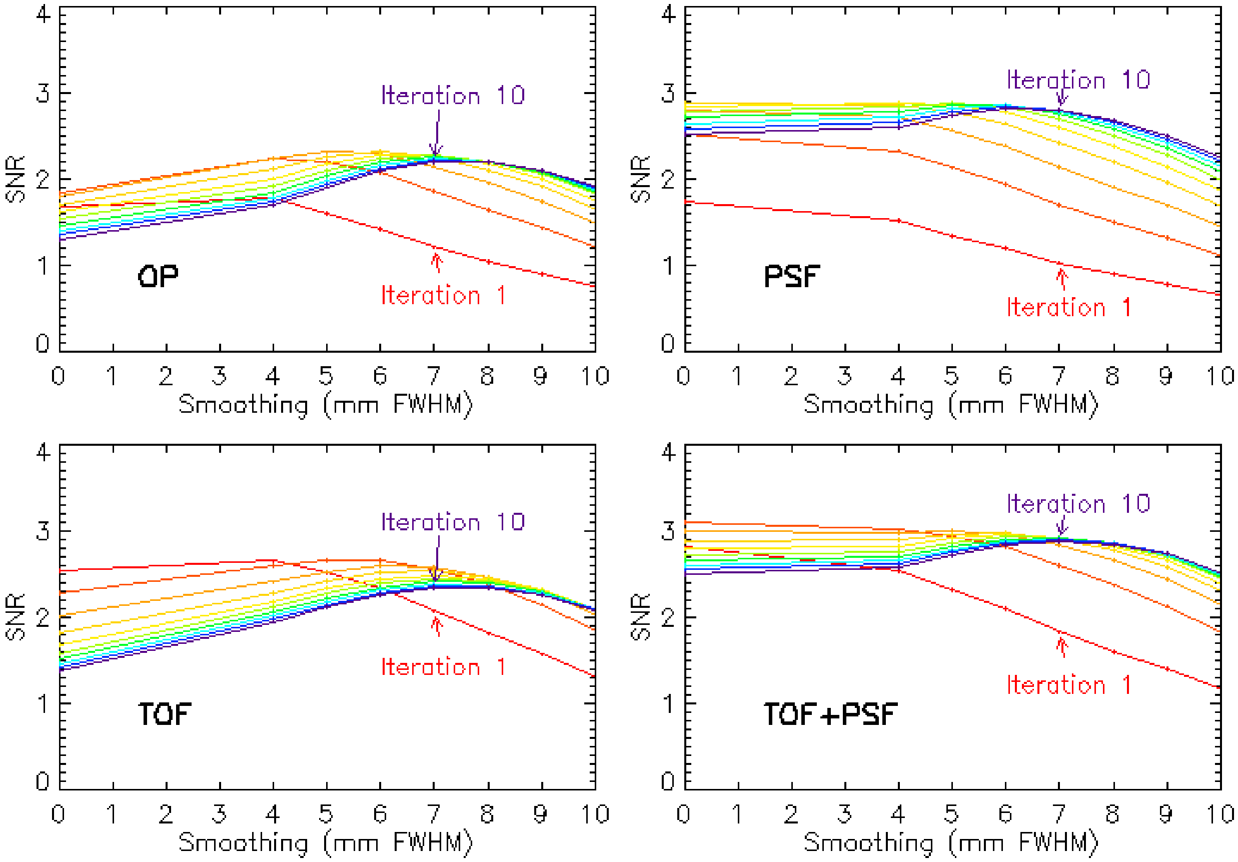
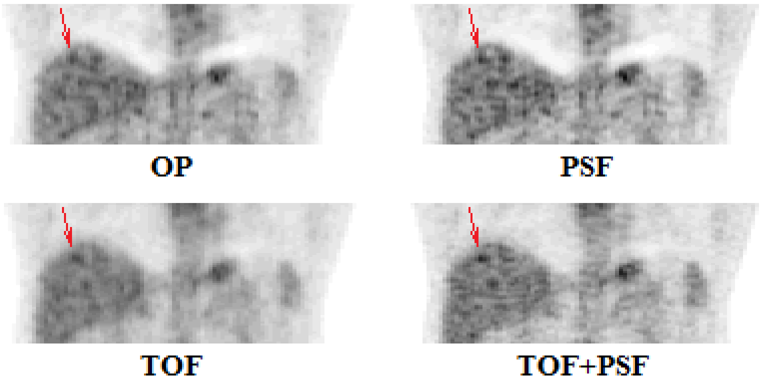


Figure 36. Plots of observer SNR; maximum SNR defined the optimal processing parameters

Based on the observer results, the reconstruction parameters were chosen as follows: OP-OSEM: 4 iterations and 6 mm filter; PSF: 5 iterations and 4 mm filter; TOF: 2 iterations and 6 mm filter; PSF+TOF: 2 iterations and no filter. The optimal reconstruction parameters vary among patients of a given population, and observer SNR was chosen to generalize these for all patients. These findings agree with those of Kadrmas et. al, where the same numerical observer was used to evaluate localization performance , instead of SNR, of lesions in an anthropomorphic phantom (Kadrmas, Casey et al. 2009). TOF reconstruction required fewer iterations, and reconstruction incorporating PSF modeling required less smoothing to achieve maximum SNR. A qualitative comparison of the four reconstructions is seen in Figure 37.



OP PSF

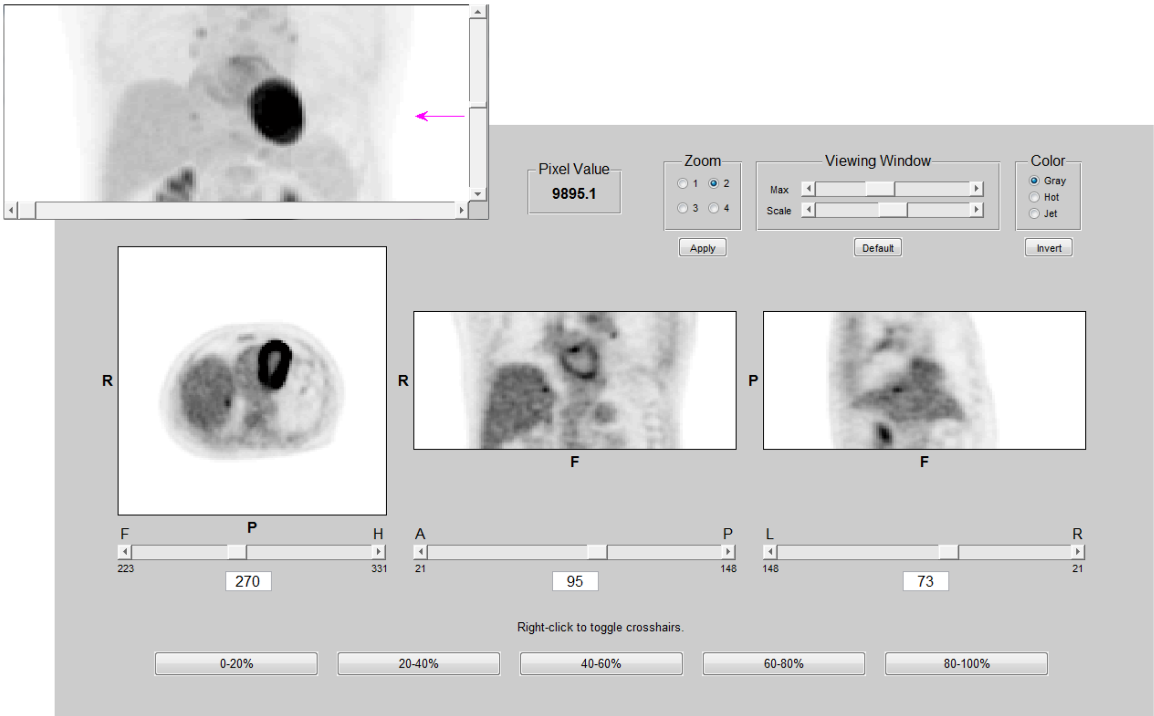
TOF TOF+PSF

Figure 37. A coronal view comparison of the four reconstructions.

Observer Studies

An experiment was designed to survey human performance in a realistic lesion detection task. A population of 33 disease-free patient bed positions was selected for the detection study. Lesion data were statistically added to these image volumes at various sites for a total of 88 lesion-present data sets. These data sets, along with the 33 lesion-absent ones, were reconstructed with the four algorithms for a study total of 484 images. In each reconstruction set, the number of lesions located at various anatomical sites was 40 in the liver, 24 in the lung cavity, 5 in the lung wall, 7 in other soft tissue regions, and 12 in the vertebrae.

A 3D image viewer was developed for this study which allowed an observer to navigate through the image volume in all orthogonal views, as well as adjust the image zoom, grayscale contrast, and color scheme. The 484 images were presented in random order to the readers, and they were instructed to examine each image, and select the most likely location of the 1-cm diameter spherical lesion, within a localization radius of 1cm. The reader had no knowledge about the presence (or absence) of the lesion and was given a 5-point scale by which to rate the confidence in their decision.



Pixel Value 9895.1

Zoom ○ 1 ● 2 ○ 3 ○ 4

Color ● Gray ○ Hot ○ Jet

Right-click to toggle crosshairs.

Figure 38. The viewer program developed to present the images and record decision data

The confidence-rated decision data were used to characterize performance and construct receiver operating characteristic (ROC) curves. This approach quantifies observer performance, in a binary task, over a large range in values of the discrimination threshold. The ROC curve plots the rate of true positives versus the rate of false positives. Localization (LROC) analysis plots the rate of correct localization versus the rate of false positives. If the data are assumed to be normally distributed, observer SNR can be calculated directly from the ROC curve.

For the analyses, an iterative, maximum-likelihood algorithm (Swensson 1996) was used to fit the corresponding probability density functions to the data from the observer tasks. This method, as well as the implementation of the localization observer, assumed the following: (1) the signal detection and localization by an observer correspond to the most likely location of the lesion in each image, (2) correct localization only occurs if the signal location is most suspicious, (3) the presence or absence of a lesion in an image has no effect on the observer's decision process, and (4) the distribution of the observer's decisions is normally distributed. These assumptions, and the consequences of their violation, are discussed further by Swensson (Swensson 1996). The physicians were encouraged to take periodic breaks within the task to minimize fatigue or strain that could affect their decision criteria.

The ROC and LROC performance of the human observers was compared to that of different 3D mathematical models. In this scheme, the model observers generated a template from a training set of images, which it then used to compute a decision statistic at every voxel (3-dimensional volume pixel) in a predefined region of each test image.

The test statistic is the numerical analog of the human perceptual measurement and is used to construct performance curves.[11] Numerical observers were evaluated in a localization task of the lesions in the liver. For each patient used in the study, the CT transmission data were used to extract this region from the PET volumes, and the segmented regions were stored for use in the localization task. This method was applied to the lungs as well, but was unsuccessful because tissue density is less correlated with tracer uptake in this region.

In this work, eight different numerical observers were evaluated: the non-prewhitening and Hotelling observers, each with 4 different channel configurations defined by Equation 5.2. As in previous works, a dense ($i \leq 10$) DOG configuration was defined by the parameters $Q = 1.67$, $\alpha = 1.4$, and $\sigma_0 = 0.005$ (Abbey and Barrett 2001), and a sparse ($i \leq 3$) DOG configuration was defined using $Q = 2$ and $\alpha = 2$. This study used 3 different frequency ranges for the sparse channels defined by $\sigma_0 = 0.010$, 0.015, and 0.020. These configurations were found to show similar performance to the human observers in the localization task.

For the human observer study, a set of images was generated solely for training purposes. The data from 16 lesions were added to 4 bed positions from 3 patients, not used in the detection task. These data sets were then reconstructed via all four reconstructions for a total of 64 images. In the training exercise, each observer had knowledge of the lesion locations. This image set (with the signal locations) was also used to train the model observers. A single template for each numerical observer was

---

[11] A numeric test statistic may not accurately represent a human observer's confidence, which is subjective and can be affected by other psychological factors.

calculated using the total population of training images. Figure 39 shows the center transaxial slice of each of the 3D templates, as well as that of the mean signal (matched filter).

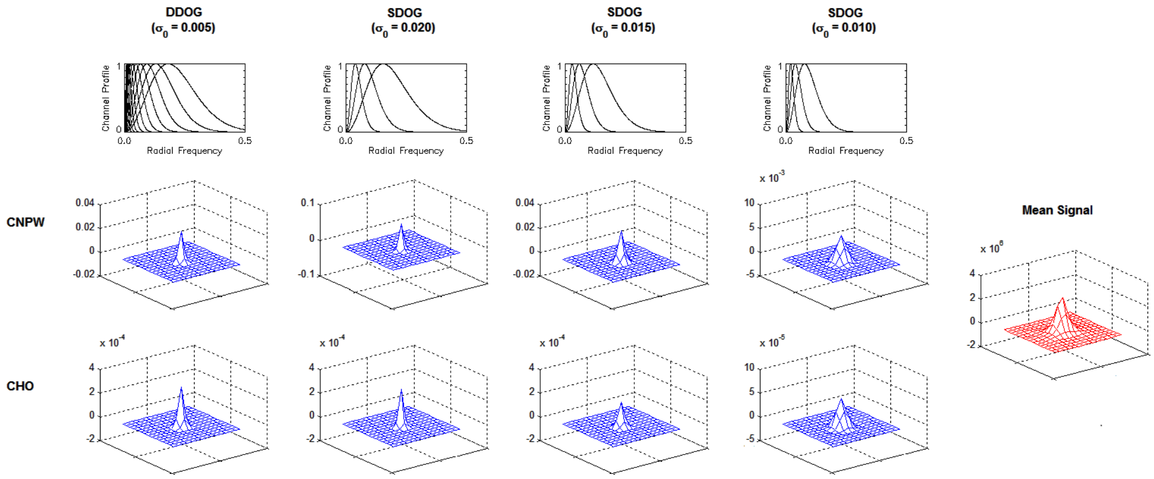


Figure 39. Templates of the 8 numerical observers evaluated in this study (blue) and the mean lesion profile (red)

The detection task of the numerical observer involved computing a test statistic at every voxel within the search area; the voxel that returned the highest statistic was then designated as the most likely location of the lesion. If the observer did not correctly locate the lesion (false positive), the decision was contributed to both the target and non-target distributions. If the observer did correctly locate the lesion, the decision contributed to the target distribution, and the location that returned the next highest statistic pooled with the non-target distribution.[12] In each task, the numerical test statistic data were collected and binned into 5 categories, to match the format to that of the human decisions.[13] Due to the range of the pixel values characteristic to different regions, the numerical test statistics belonging to different organs were not directly comparable. Thus, performance of the numerical observer was evaluated within the liver and lungs independently.

Results

The ROC and LROC performance curves of the 3 human observers for all lesions in all anatomical locations are shown in Figure 40.

---

[12] This approach is valid under the assumption that the presence (or absence) of a lesion does not affect the observer's decision making process.

[13] The data were assumed to be normally distributed with the bin widths defined by the standard deviations of the sets.

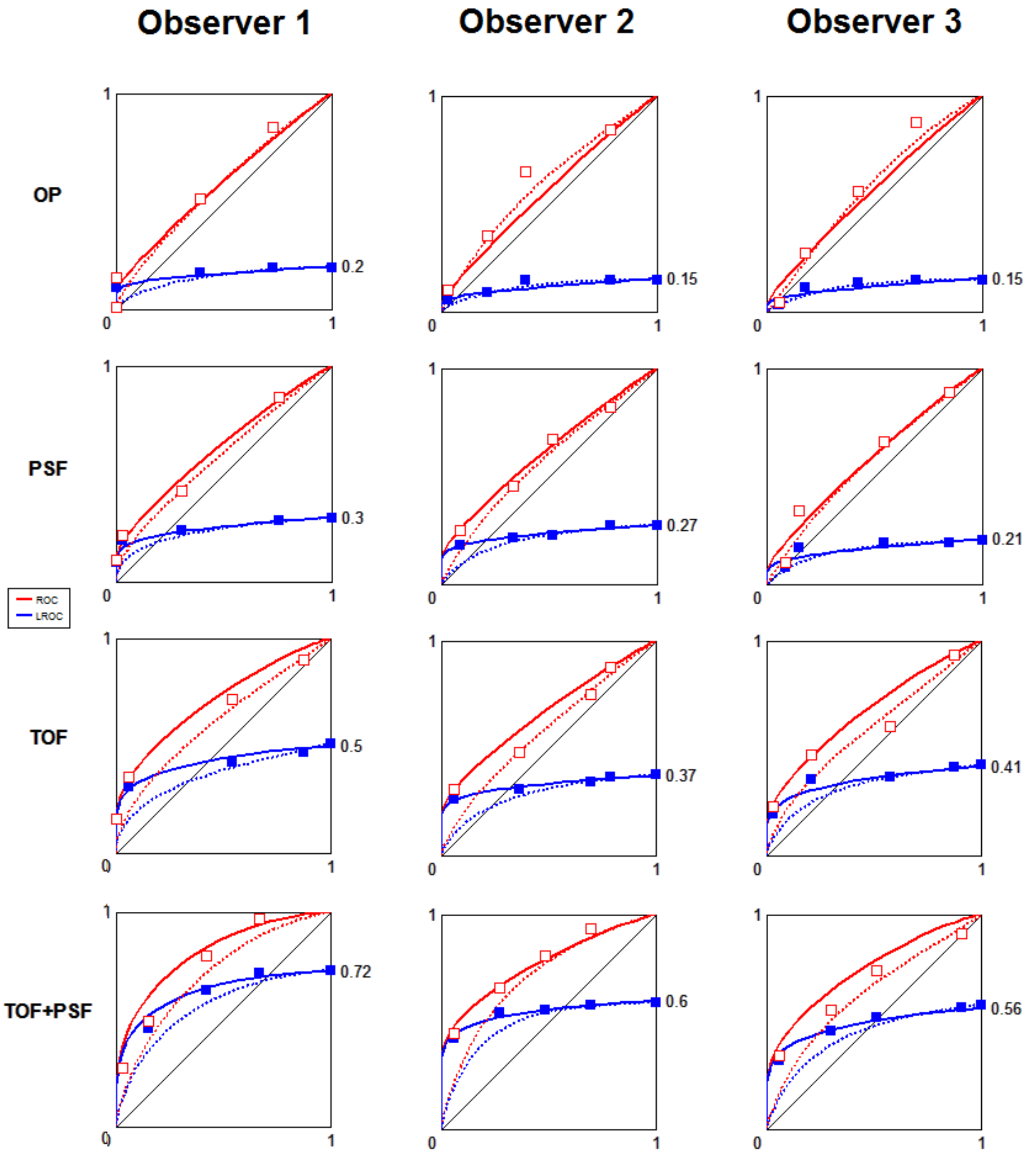


Figure 40. Performances of 3 physician observers in the detection and localization task of all lesions in all anatomical locations

The corresponding mean performances of are seen in Figure 41.

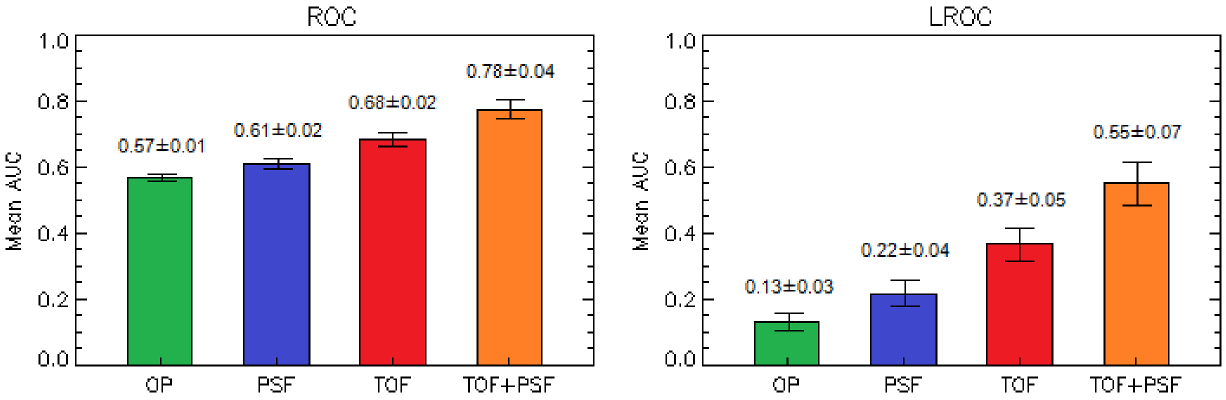


Figure 41. Mean performances of the physicians in the detection and localization task

The results from the human study (AUC_{ROC}; AUC_{LROC}), averaged over the 3 observers, were: OP (0.57±0.01; 0.13±0.03), PSF (0.61±0.02; 0.22±0.04), TOF (0.68±0.02; 0.37±0.05), and TOF+PSF (0.78±0.04; 0.55±0.07). The overall probabilities of correct localization were OP (0.17±0.03), PSF (0.26±0.03), TOF (0.43±0.08), and TOF+PSF (0.63±0.13).

There were significant differences in the performances of the human observers corresponding to lesion SNR at different anatomical sites. Figure 42 shows the localization curves for lesions located in the liver and lungs. The improved performances within the lung lesions are attributed to higher lesion SNR in this region.

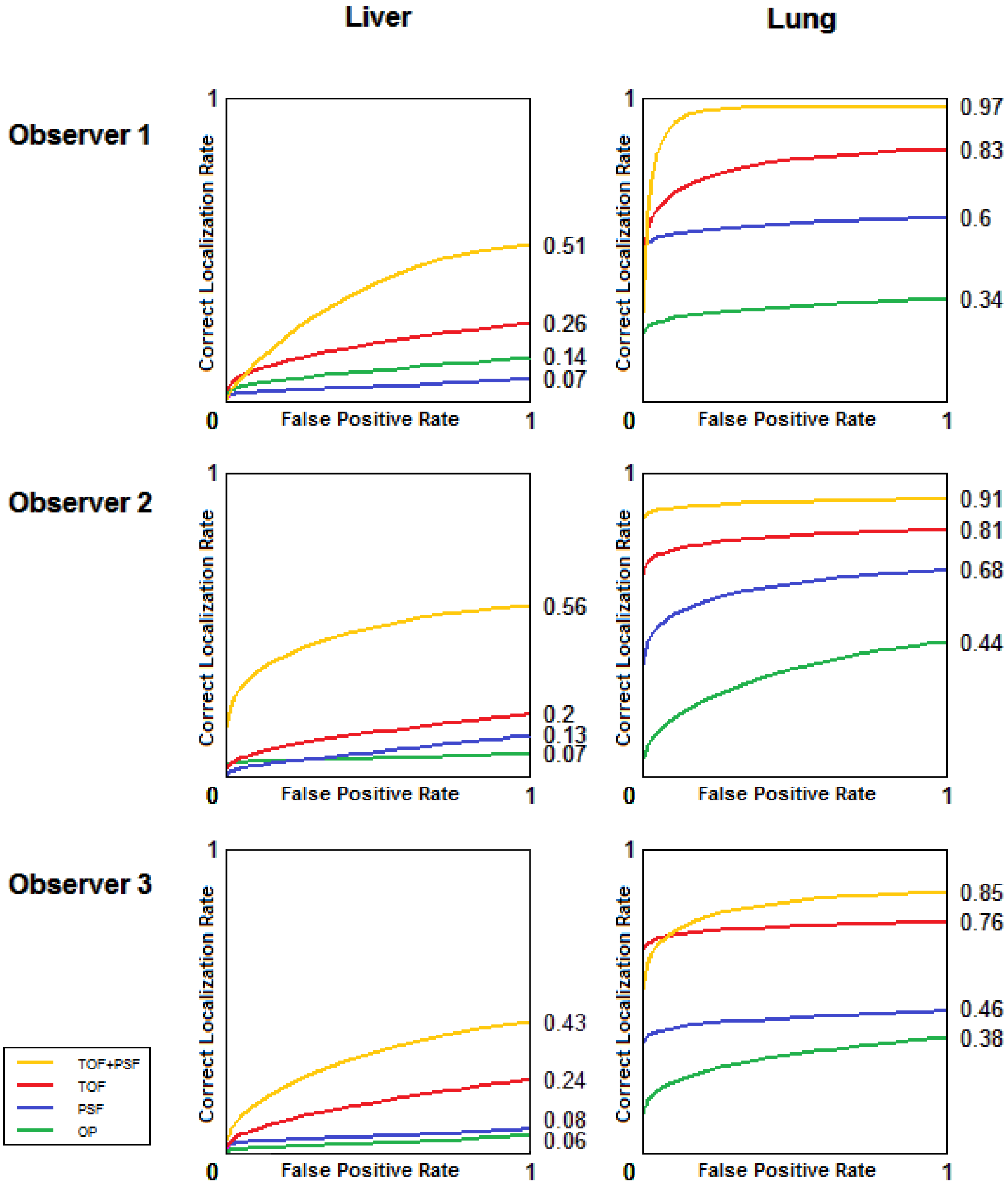


Figure 42. Comparison of physician performance in localization of liver and lung lesions

The performances of the human observers within the population of liver lesions are seen in Figure 43, separated by patient BMI. In some cases, fitting the individual class probabilities independently (dotted lines) resulted in observer performance falling below zero, exhibiting upward hook deflections, or ROC curves that drop below the unity line. These are examples of impossible performance curves, as mentioned in chapter 3. The parametric model (solid lines) was generally better suited to fit these data.

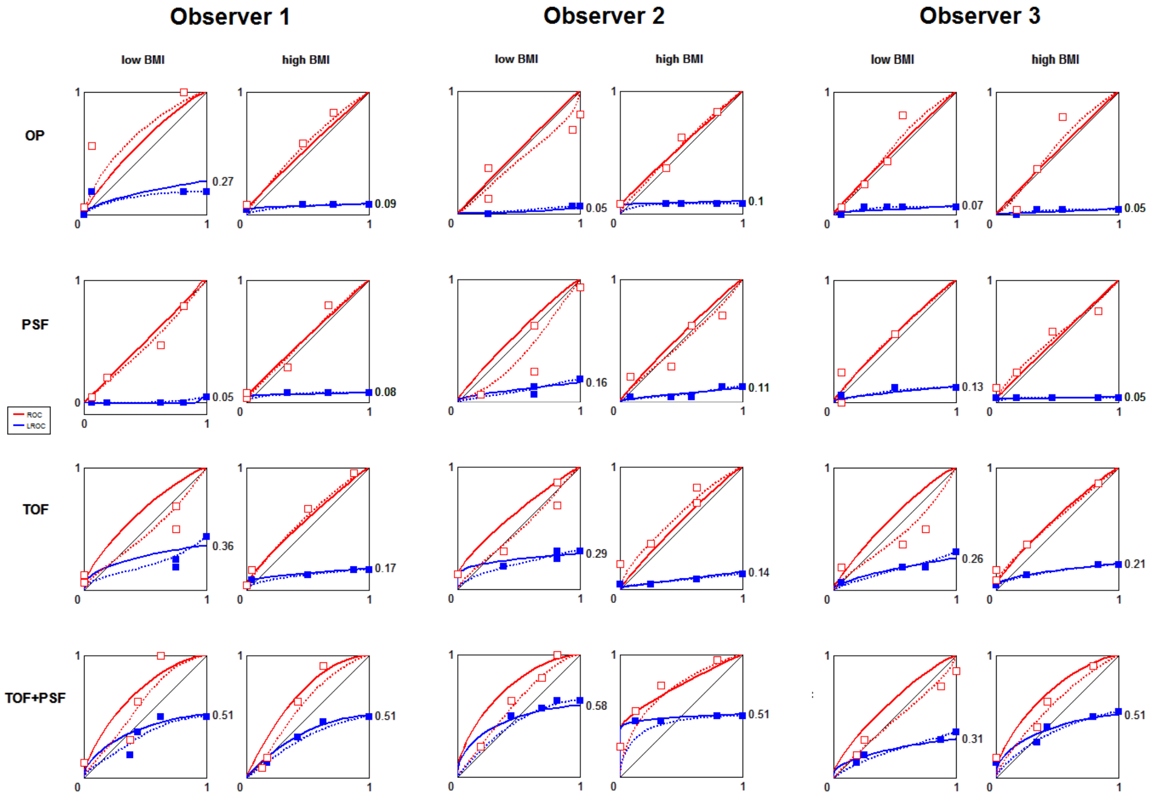


Figure 43. Human observer performance in the detection and localization task of lesions located in the liver, grouped by patient BMI

The performance of the human observers was compared to that of each of the numerical observers, in the same localization task of lesion in the liver. The LROC of each is seen in Figure 44.

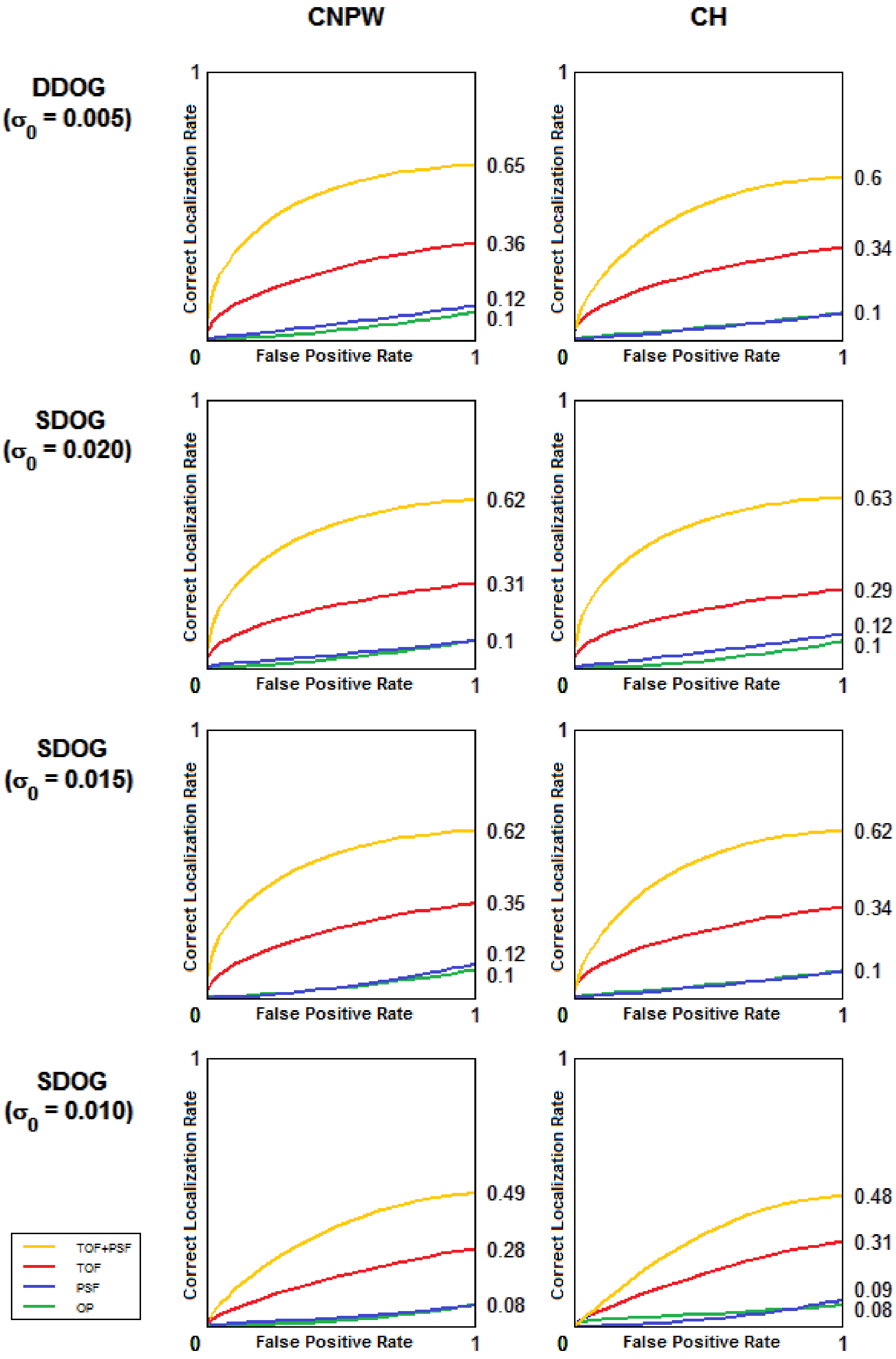


Figure 44. Comparison of the numerical observer models in localization of liver lesions

The filter selection of the numerical observer had a greater impact on the performance than the whitening operation.

As seen in Figure 45, good agreement was found between the mean performance of human and mathematical observers in the localization task of lesions in the liver.

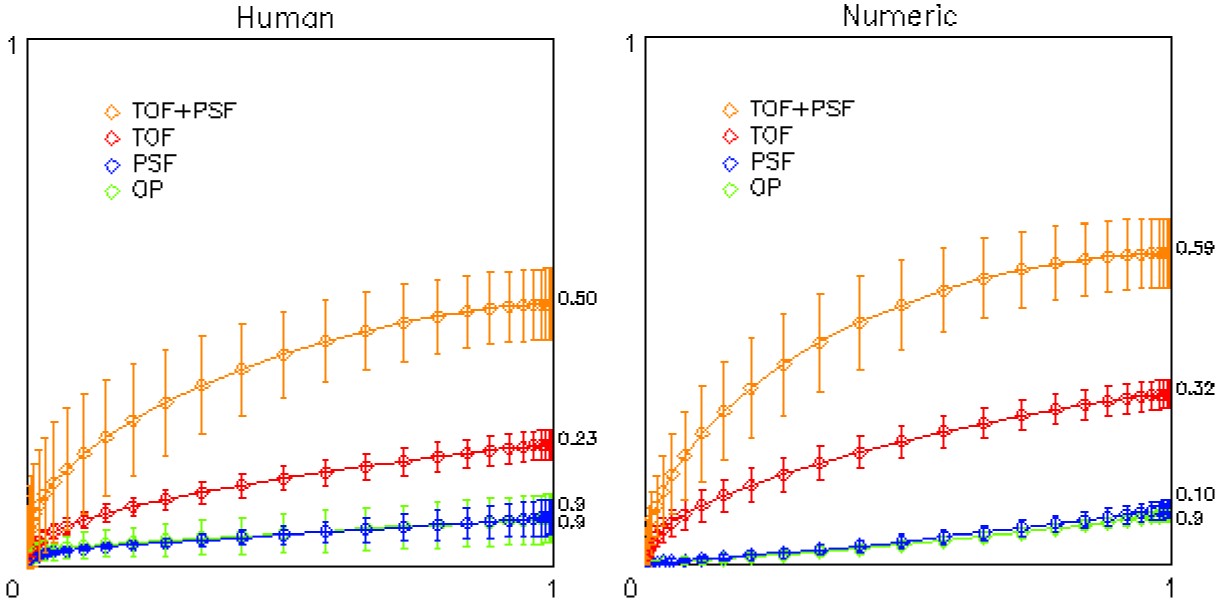


Figure 45. LROC curves showing mean localization performance of liver lesions for 3 human observers (left) and 8 numerical models (right)

The 8 numerical observers were used to evaluate the relative benefit of TOF+ PSF in the clinical data for different patient sizes, lesion contrasts, and scan times. Figure 46 shows the observer SNR for different lesion intensity and scan time. Each reconstruction set was processed with the number of iterations found to maximize observer SNR (Figure 36).

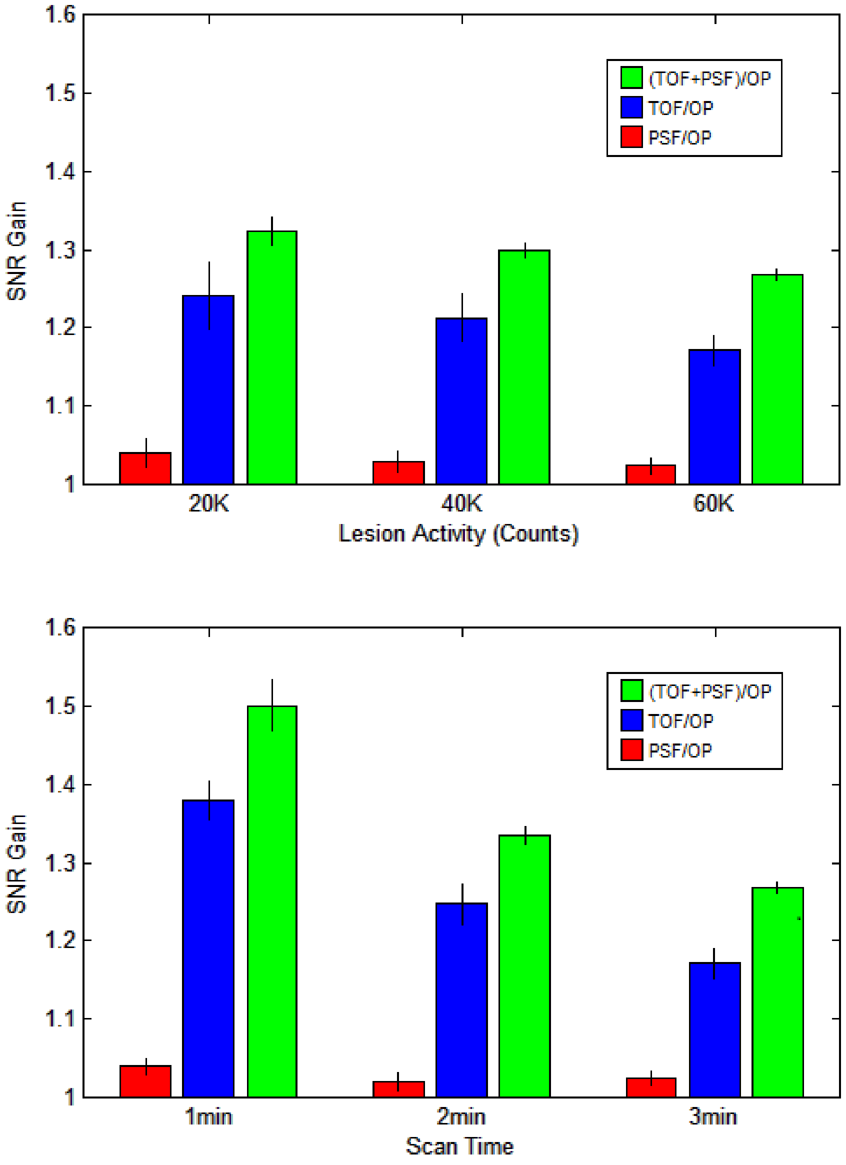


Figure 46. Plots showing mean improvement in observer SNR of PSF, TOF, and TOF+PSF compared to OP, for various lesion contrasts and acquisition times.

The data in Figure 46a are from the 3-minute scans and Figure 46b shows the data incorporating lesion with 60K counts. The greatest improvement is clearly seen in the faintest lesions and in the shortest scans.

Figure 47 shows the gain in observer performance for patients with low and high BMI.

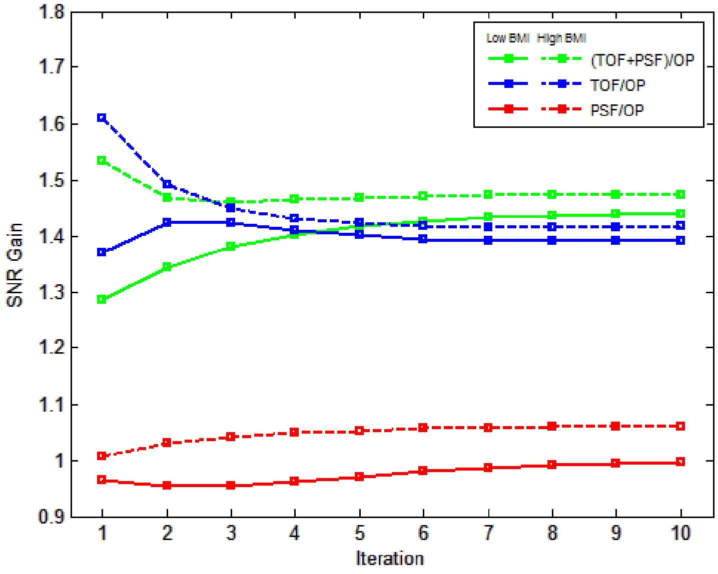


Figure 47. Mean observer SNR improvement for different patient sizes.

TOF showed more improvement than PSF. For all iterations of each respective algorithm, the gain of PSF, TOF, and TOF+PSF was higher for the larger subjects.

Discussion

A lesion detection study has been developed to assess the clinical benefits of PET reconstructions incorporating PSF and TOF. The use of real patient images with simulated lesions depicted an actual clinical task, that is, the detection and localization of small lesions with focal uptake in the liver, lungs, or other tissues. Numerical observer models were used to optimize experimental parameters as well as predict human performance. The assistance of several board-certified physicians provided validation of our numerical models.

The experiment was not designed to evaluate the observer's ability to detect disease. Rather it was designed to survey the ability to detect a predefined signal in an anatomical background. In a forced detection task, for any observer and random set of *lesion-absent* images, there is a given rate of detections (false-positive rate). These decisions, whether or not they are accurate clinically, serve to define a baseline for the evaluation. A study like this measures the characteristic differences when "known lesions" are introduced into the image population and cause the observer detection rate to rise above the baseline. This should necessarily restrict the ROC performance curve from deflecting below the unity line. The images presented to the physicians included a range of lesion contrasts, and many were very low intensity, making the detection task challenging; A_{ROC} was close to 0.5 for the baseline OP reconstruction (essentially the same as a flipping a coin). This enhanced the gains in performance between the different reconstructions and emphasized the clinical benefits for the detection of small metastatic lesions that are "just visible". The performance of the physicians demonstrated that TOF+PSF reconstruction yielded superior performance over PSF or TOF alone, supporting the results found by the mathematical observers for the same localization task. This study simulated an actual clinical task but certainly did not account for all the complications of the diagnostic process.

Mathematical observers base their decisions only on the information that is available through training. The models are conditioned under the presumption that each element of the sample is part of the same random. This assumption is rarely a good fit with real clinical imaging applications, where the amount of available training information is often insufficient to outweigh the randomness within the data. In this case, the ensemble cannot be completely characterized, i.e. a full-rank covariance matrix cannot be constructed. This can be overcome through the use of limited, sub-images containing the important features and frequency-selective filtering. Both methods reduce the dimensionality of the problem and require prior information about the signal location.

Contradictory results have been reported on the ability of a human observer to whiten data in signal detection tasks (Myers, Barrett et al. 1985; Rolland and Barrett 1992). The Hotelling observers in this study had internal noise added to the whitening operation to account for human error, and we found good agreement in these performances and those of the CNPW observers. The frequency selection of the channelized filter had more influence on observer performance than the whitening operation.

Noise in the PET data had a large impact on the reconstruction process, as seen in Figure 47. In iterative reconstruction, noise in the data obstructs the convergence of the algorithm and then causes it to diverge. After too many iterations, the signal contrast has already converged, and SNR decreases because image noise increases. This effect is accelerated when using TOF. For large patients, with very noisy data, the accelerated rate of TOF provided the largest SNR gain at the earliest iterations. In smaller patients, with less noisy data, the baseline reconstruction performed relatively better at the first iteration, and the maximum gain associated with TOF peaked later. Further, the convergence rate was delayed by incorporation of the PSF model. In small patients, resolution modeling offered little benefit and actually degraded observer performance in this task; this was also seen in the performance of 2 of the 3 human observers.

The analyses by our models showed that both TOF and PSF offer greater overall benefit for low-intensity lesions, shorter acquisition times, and in larger patients, consistent with the findings of previous groups (Armstrong, Williams et al. 2011; El Fakhri, Surti et al. 2011) The resolution modeling of the PSF algorithm produces higher contrast values for small signals, which can increase numerical observer SNR in certain tasks, provided the algorithm has sufficiently converged; at lower iterations however, the PSF reconstruction may yield suboptimal performance. The noise reduction across the FOV from the TOF information leads to better performance in localization tasks, for human and mathematical observers. In every case, the reconstruction with the combination of both PSF and TOF outperformed all other reconstructions.

Chapter 6

Conclusions and Future Prospects

Summary of Dissertation

The aim of this dissertation was to evaluate the effect of recent PET reconstruction techniques on image quality, in the context of a defined clinical task. This work has combined physics investigations with current PET technology and clinical application. Several techniques were used to accomplish the goals of this research. These included: phantom experiments, resampling of list-mode data, sinogram manipulation, analysis of PET image variance, image filtering and Fourier processing, probability function modeling, and linear discriminant analysis. This study required extensive computational development of several tools, necessary to frame it as a signal detection experiment, applied to a realistic problem. These included: scripts to reconstruct thousands of images with adaptive parameters, the image viewing platform for the physicians, various numerical observer models, and the automated quantification of performance in detection and localization tasks. These tools were developed in various languages including C, IDL, and MATLAB.

The dissertation began with an introduction to PET, including its origin and its evolution as a clinical tool. The underlying physics of the modality followed in Chapter 2, including the statistical corrections necessary to accurately quantize PET images. The two reconstruction techniques investigated in this work, time-of-flight and point spread function modeling, were introduced at the end of the chapter. These algorithms were evaluated by various observers in a signal detection task, and Chapter 3 covered the basis of signal detection theory. The receiver operating characteristic and localization receiver operating characteristic methodologies, used to quantify observer performance, were discussed in this chapter, along with their foundations in the model of the psychophysical response of an observer. The chapter concluded with an explanation of the maximum likelihood algorithm used to estimate density functions from the observer data. Chapter 4 continued by discussing linear mathematical observer models and their performance relative to human observers. Two numerical schemas were used in this work, non-prewhitening and Hotelling. NPW used only the first-order moments of the observer statistics to define the linear discriminant while the Hotelling also took into account the second-order properties of the statistical distributions. Channelized filters, used to extract the important signal information from the images, were reviewed at the end of the chapter.

The main topic and description of the dissertation work was covered in Chapter 5: a rigorous investigation of lesion detection and localization performance of PSF and TOF reconstructions in clinical PET. The chapter addressed previous work in the field and noted the lack of quantified clinical evaluation, giving the motivation for the current study. Phantom scans and clinical PET data were used to simulate small, focal disease in otherwise normal images. Four iterative reconstructions were used, and each scheme was optimized based on numerical observer SNR. Images reconstructed by these parameters were presented to several certified physicians, and their performances were quantified by ROC and LROC analyses. Several numerical observers were subjected to the same localization task, and these results were compared to those of the human observers.

This work is the first study to evaluate PSF and TOF in a simulated clinical task, while comparing the performance of experienced nuclear medicine physicians to that of an array of numerical observer models. The investigations presented here were conducted independently, and many of the results show strong agreement with those of related works by other groups. The analyses show a large benefit for lesion detection and localization when using the combination of TOF and PSF simultaneously.

Research Contributions

Modern clinical tools are able to assist doctors in unprecedented ways; the evolution of PET image reconstruction, from the early analytic methods to the recent algorithms covered here, is indicative of the overall progress in the molecular imaging field. Ongoing translational work is necessary to evaluate the benefits and limitations of new developments in the medical setting, in this case TOF and PSF. These reconstructions are relatively new and not routinely used in most clinical institutions. This work analyzed the image quality of these techniques in the context of the physician's ability to differentiate diseased from normal tissue. The results presented here suggest that patient management can be significantly improved, and these data may help support changes to current image processing procedures.

At the core of this study lies a signal detection experiment. In this respect, the value of this research is twofold: to evaluate the performance of experienced human observers, and to cross-validate these findings with those from various numerical models. It is difficult to reduce the human observers' response to stimuli to a single metric, such as observer SNR. However, with the proper constraints, this process can be effectively surveyed, characterized, and modeled. In the medical setting, computer observers are designed to predict human performance when large amounts of data need to be processed. These numerical models are generally divided into 2 classes, based on the presumed ability (or inability) of the human observer to whiten correlated noise within the test images. Usually, a study will implement one type for all analyses; the current study investigated both, using several configurations of the channelized filters in each. The data are valuable, as they show a comparison of numerical observer performances in a realistic localization task. Moreover, this evaluation showed that the selection of the filter was more important than the inclusion of the whitening operation. In the future, this work may serve as a reference for researchers interested in choosing an appropriate numerical model for their observer tasks.

The tools developed for this study will be used in future studies as well. The MATLAB viewing program can be adapted to present images and record decision data for any human observer study. This program has been extended to include FROC mode, which allows multiple viewer choices within each image. The same numerical observer models designed for this research are now being used to investigate the performances of various PET reconstructions in the detection of myocardial defects. The code used for anatomical organ segmentation will also be useful in future work.

PET in the Future

Despite its relatively brief lifetime, clinical PET is an indispensable diagnostic tool, used in many branches of medicine. Naturally, its integration into clinical practice will only increase with time and technological improvements.

The addition of the CT to PET, creating the new modality PET/CT, marked one of the most significant developments in the field. This had a huge impact on reimbursement policies and was arguably the largest catalyst for clinical integration. Various new hybrid modalities will be spawned following this model, just as SPECT/CT was; PET/MR is one such recently developed modality. While its routine applications have yet to be established, this sophisticated instrument has potential to be a very powerful clinical tool.

The transition from linear to iterative reconstructions in the last decade significantly enhanced image quality. New methods to correct motion, scatter, and other sources of noise from the acquisition process are being investigated and will further benefit the quality of PET images. Incorporation of TOF measurements is one of the latest major advancements, and as detector performance advances, performance will continually improve. The electronics on today's PET systems can achieve timing resolution of 78 picoseconds. However, since current detector hardware is based on scintillating crystals coupled to photomultiplier tubes, the conversion of gamma energy to light, and then to electronic pulse, degrades this resolution by a factor of about 7. The detection systems will see a transition to solid-state detectors, improving energy and timing resolutions. Although a distant goal, in the future a PET system may be able to pinpoint the location of every annihilation, directly placing each registered count in its respective pixel, eliminating the reconstruction process altogether. This would allow the spatial resolution to approach the intrinsic limit of PET, which is currently restricted (in clinical systems) by the acolinearity of the coincident photons.

Furthermore, the current paradigm of reconstructing PET data to produce static images may also change in the future. Fully quantitative methods such as dynamic PET and tracer kinetic analysis are capable of providing more valuable information of physiological processes than conventional “snap shots” of tracer uptake. However, these techniques are limited in the clinic by the need for long scan times, limited reconstruction performance, and technologically cumbersome procedures. Many groups are investigating reconstructions that directly generate parameter images from the raw PET data. These parameterized images show a complete map of tracer rate constants for every pixel, and could provide more physiologically relevant information to physicians.

The most important advances to emission tomography, however, may not lie in the tomographs, but rather in the development of new biomarkers. FDG is by far the most commonly used tracer in PET because it tracks cellular metabolism, and is sensitive to many pathologic processes. This non-specificity is its greatest strength, but is also its greatest weakness, as it reveals healthy, energy-consuming tissue as well. New tracers are being developed to take advantage of properties unique to individual pathology. A current example is in the field of neurology, which has been gaining much recent attention for PET for the early diagnoses of Alzheimer’s disease. This is becoming possible by exploiting the neocortical binding of newly synthesized compounds to amyloid plaque deposits. In the coming years, nuclear medicine will see many new radiotracers emerge, specifically tailored to an increasing number of diseases.

List of References

Abbey, C. and H. Barrett (2001). "Human-and model-observer performance in ramp-spectrum noise: effects of regularization and object variability." JOSA A **18**(3): 473-488.

Abramowitz, M. and I. Stegun (1972). Bernoulli and Eulear Polynomials and the Euler-Maclaurin Formula, $23.1 in Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing, New York: Dover.

Allemand, R., C. Gresset, et al. (1980). "Potential advantages of a cesium fluoride scintillator for a time-of-flight positron camera." Journal of Nuclear Medicine **21**(2): 153.

Armstrong, I. S., H. A. Williams, et al. (2011). Accuracy and variability of quantitative measurements using PET with time-of-flight information and resolution modelling. Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), IEEE.

Bailey, D. and S. Meikle (1994). "A convolution-subtraction scatter correction method for 3D PET." Physics in Medicine and Biology **39**: 411.

Barney, J., J. Rogers, et al. (1991). "Object shape dependent scatter simulations for PET." Nuclear Science, IEEE Transactions on **38**(2): 719-725.

Barrett, H., J. Yao, et al. (1993). "Model observers for assessment of image quality." Proceedings of the National Academy of Sciences of the United States of America **90**(21): 9758-9765.

Barrett, H. H., T. Gooley, et al. (1992). "Linear discriminants and image quality." Image and Vision Computing **10**(6): 451-460.

Barrett, H. H. and K. J. Myers (2004). Foundations of image science. Hoboken, NJ, Wiley.

Bendriem, B. and D. W. Townsend (1998). The theory and practice of 3D PET, Springer.

Bergström, M., L. Eriksson, et al. (1983). "Correction for scattered radiation in a ring detector positron camera by integral transformation of the projections." Journal of Computer Assisted Tomography **7**(1): 42.

Beyer, T., P. Kinahan, et al. (1994). The use of X-ray CT for attenuation correction of PET data, IEEE.

Beyer, T., D. Townsend, et al. (2000). "A combined PET/CT scanner for clinical oncology." Journal of Nuclear Medicine **41**(8): 1369.

Bietendorf, J. (2004). "FDG PET Reimbursement." Journal of Nuclear Medicine Technology **32**(1): 33.

Brownell, G. and W. Sweet (1953). "Localization of brain tumors with positron emitters." Nucleonics **11**(11): 40-45.

Budinger, T. (1983). "Time-of-flight positron emission tomography: status relative to conventional PET." Journal of Nuclear Medicine **24**(1): 73-78.

Burgess, A. (1985). "Effect of quantization noise on visual signal detection in noisy images." JOSA A **2**(9): 1424-1428.

Burgess, A. and B. Colborne (1988). "Visual signal detection. IV. Observer inconsistency." Journal of the Optical Society of America A **5**(4): 617-627.

Burgess, A., R. Shaw, et al. (1999). "Noise in imaging systems and human vision." JOSA A **16**(3): 618.

Campbell, F. and J. Robson (1968). "Application of Fourier analysis to the visibility of gratings." The Journal of Physiology **197**(3): 551.

Carson, R., M. Daube-Witherspoon, et al. (1988). "A method for postinjection PET transmission measurements with a rotating source." Journal of Nuclear Medicine **29**(9): 1558.

Casey, M. (2007). "Point spread function reconstruction in PET." Siemens Medical Solution, Knoxville, USA.

Casey, M. (2008). "Improving PET With HD•PET + Time of Flight." Siemens Medical Solution, Knoxville, USA.

Casey, M. and R. Nutt (1986). "A multicrystal two dimensional BGO detector system for positron emission tomography." IEEE Trans. Nucl. Sci **33**(1): 460-463.

Chan, H., K. Doi, et al. (1990). "Improvement in radiologists' detection of clustered microcalcifications on mammograms: the potential of computer-aided diagnosis." Investigative Radiology **25**(10): 1102.

Cherry, S., M. Dahlbom, et al. (1991). "3D PET using a conventional multislice tomograph without septa." Journal of Computer Assisted Tomography **15**(4): 655.

Conti, M., B. Bendriem, et al. (2005). "First experimental results of time-of-flight reconstruction on an LSO PET scanner." Physics in Medicine and Biology **50**: 4507-4526.

Daube-Witherspoon, M. E., S. Matej, et al. (2011). Impact of resolution modeling on accuracy and precision of lesion contrast measurements. Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), IEEE.

Doi, K., M. Giger, et al. (1992). "Computer-aided diagnosis: development of automated schemes for quantitative analysis of radiographic images." Semin Ultrasound CT MR **13**(2): 140–152.

El Fakhri, G., P. A. Santos, et al. (2007). "Impact of Acquisition Geometry, Image Processing, and Patient Size on Lesion Detection in Whole-Body 18F-FDG PET." Journal of Nuclear Medicine **48**(12): 1951-1960.

El Fakhri, G., S. Surti, et al. (2011). "Improvement in lesion detection with whole-body oncologic time-of-flight PET." Journal of Nuclear Medicine **52**(3): 347-353.

Fiete, R., H. Barrett, et al. (1987). "Hotelling trace criterion and its correlation with human-observer performance." JOSA A **4**(5): 945-953.

Fisher, R. (1936). "The use of multiple measures in taxonomic problems." Ann. Eugenics **7**: 179-188.

Gifford, H., P. Kinahan, et al. (2007). "Evaluation of multiclass model observers in PET LROC studies." IEEE transactions on nuclear science **54**(1 Part 1): 116-123.

Gifford, H., M. King, et al. (2000). "Channelized Hotelling and human observer correlation for lesion detection in hepatic SPECT imaging." Journal of Nuclear Medicine **41**(3): 514-521.

Gifford, H., M. King, et al. (2005). "A comparison of human and model observers in multislice LROC studies." IEEE transactions on medical imaging **24**(2): 160-169.

Giger, M. L. and K. Doi (1987). "Effect of pixel size on detectability of low-contrast signals in digital radiography." JOSA A **4**(5): 966-975.

Grootoonk, S., T. Spinks, et al. (1996). "Correction for scatter in 3D brain PET using a dual energy window method." Physics in Medicine and Biology **41**: 2757.

Hoffman, E., M. Phelps, et al. (1976). "Design and performance characteristics of a whole-body positron transaxial tomograph." Journal of Nuclear Medicine **17**(6): 493.

Hotelling, H. (1931). "The generalization of Student's ratio." The Annals of Mathematical Statistics: 360-378.

Hounsfield, G. (1973). "Computerized transverse axial scanning (tomography): Part 1. Description of system." British Journal of Radiology **46**(552): 1016.

Hubbell, J. and U. S. N. B. o. Standards (1969). Photon cross sections, attenuation coefficients, and energy absorption coefficients from 10 keV to 100 GeV, US National Bureau of Standards; for sale by the Supt. of Docs., US Govt. Print. Off.

Jakoby, B., Y. Bercier, et al. (2008). "Performance investigation of a time-of-flight PET/CT scanner."

Judy, P. F., R. G. Swensson, et al. (1981). "Lesion detection and signal–to–noise ratio in CT images." Medical Physics **8**: 13.

Kadrmas, D. J., M. E. Casey, et al. (2009). "Experimental comparison of lesion detectability for four fully-3D PET reconstruction schemes." IEEE Trans Med Imaging **28**(4): 523-534.

Kadrmas, D. J., M. E. Casey, et al. (2009). "Impact of Time-of-Flight on PET Tumor Detection." Journal of Nuclear Medicine **50**(8): 1315-1323.

Kadrmas, D. J., M. B. Oktay, et al. (2012). "Effect of Scan Time on Oncologic Lesion Detection in Whole-Body PET." Nuclear Science, IEEE Transactions on **59**(5): 1940-1947.

Karp, J. S., S. Surti, et al. (2008). "Benefit of Time-of-Flight in PET: Experimental and Clinical Results." Journal of Nuclear Medicine **49**(3): 462-470.

Karshenboim, S. G. (2003). "Precision study of positronium: testing bound state QED theory." Arxiv preprint hep-ph/0310099.

Kinahan, P., D. Townsend, et al. (1998). "Attenuation correction for a combined 3D PET/CT scanner." Medical Physics **25**: 2046.

Lois, C., B. W. Jakoby, et al. (2010). "An Assessment of the Impact of Incorporating Time-of-Flight Information into Clinical PET/CT Imaging." Journal of Nuclear Medicine **51**(2): 237-245.

Loo, L., K. Doi, et al. (1984). "A comparison of physical image quality indices and observer performance in the radiographic detection of nylon beads." Physics in Medicine and Biology **29**: 837.

Melcher, C. and J. Schweitzer (1992). "A promising new scintillator: cerium-doped lutetium oxyorthosilicate." Nuclear Instruments and Methods in Physics Research

Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **314**(1): 212-214.

Michel, C., M. Sibomana, et al. (1998). Preserving Poisson characteristics of PET data with weighted OSEM reconstruction. Nuclear Science Symposium, 1998. Conference Record. 1998 IEEE, IEEE.

Mullani, N., J. Markham, et al. (1980). "Feasibility of time-of-flight reconstruction in positron emission tomography." Journal of Nuclear Medicine **21**(11): 1095.

Mumcuoglu, E., R. Leahy, et al. (1996). "Bayesian reconstruction of PET images: methodology and performance analysis." Physics in Medicine and Biology **41**: 1777.

Myers, K., H. Barrett, et al. (1985). "Effect of noise correlation on detectability of disk signals in medical imaging." Journal of the Optical Society of America A **2**(10): 1752-1759.

Myers, K. J. and H. H. Barrett (1987). "Addition of a channel mechanism to the ideal-observer model." JOSA A **4**(12): 2447-2457.

Ollinger, J. (1996). "Model-based scatter correction for fully 3D PET." Physics in Medicine and Biology **41**: 153.

Panin, V. Y., F. Kehren, et al. (2006). "Fully 3-D PET reconstruction with system matrix derived from point source measurements." Medical Imaging, IEEE Transactions on **25**(7): 907-921.

Patlak, C., R. Blasberg, et al. (1983). "Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data." J Cereb Blood Flow Metab **3**(1): 1-7.

Phelps, M., K. E., et al. (1981). "Metabolic mapping of the brain's response to visual stimulation: studies in man." Science **211**: 1445-1448.

Phelps, M., E. Hoffman, et al. (1976). "Tomographic images of blood pool and perfusion in brain and heart." Journal of Nuclear Medicine **17**(7): 603.

Phelps, M., E. Hoffman, et al. (1975). "Application of annihilation coincidence detection to transaxial reconstruction tomography." Journal of Nuclear Medicine **16**(3): 210.

Phelps, M., S. Huang, et al. (1979). "Tomographic measurement of local cerebral glucose metabolic rate in humans with (F 18) 2 fluoro 2 deoxy D glucose: validation of method." Annals of Neurology **6**(5): 371-388.

Rolland, J. and H. Barrett (1992). "Effect of random background inhomogeneity on observer detection performance." Journal of the Optical Society of America A **9**(5): 649-658.

Schaefferkoetter, J., M. Casey, et al. (2013). "Clinical impact of time-of-flight and point response modeling in PET reconstructions: a lesion detection study." Physics in medicine and biology **58**(5): 1465.

Schlyer, D. (2004). "PET tracers and radiochemistry." ANNALS-ACADEMY OF MEDICINE SINGAPORE **33**(2): 146-154.

Shapiro, J. (2002). Radiation protection : a guide for scientists, regulators, and physicians. Cambridge, Mass. [u.a.], Harvard University Press.

Shepp, L. and Y. Vardi (1982). "Maximum likelihood estimation for emission tomography." IEEE Trans. Med. Imaging **1**: 113–121.

Shreve, P. and D. W. Townsend (2008). Clinical PET-CT. New York; London, Springer.

Smith, W. E. and H. H. Barrett (1986). "Hotelling trace criterion as a figure of merit for the optimization of imaging systems." JOSA A **3**(5): 717-725.

Stearns, C. (1995). "Scatter correction method for 3D PET using 2D fitted Gaussian functions." J Nucl Med **36**: 105.

Surti, S. and J. S. Karp (2009). "Experimental evaluation of a simple lesion detection task with time-of-flight PET." Physics in Medicine and Biology **54**(2): 373-384.

Surti, S., J. Scheuermann, et al. (2011). "Impact of Time-of-Flight PET on Whole-Body Oncologic Studies: A Human Observer Lesion Detection and Localization Study." Journal of Nuclear Medicine **52**(5): 712-719.

Swensson, R. (1996). "Unified measurement of observer performance in detecting and localizing target objects on images." Medical Physics **23**: 1709-1725.

Tarantola, G., F. Zito, et al. (2003). "PET instrumentation and reconstruction algorithms in whole-body applications." Journal of Nuclear Medicine **44**(5): 756.

Tomitani, T. and E. Tanaka (1980). Noise Characteristics of Positron CT Using Time of Flight. Third Symposium on Physical and Technical Aspects of Transmission and Emission Computed Tomography.

Tong, S., A. Alessio, et al. (2010). "Noise and signal properties in PSF-based fully 3D PET image reconstruction: an experimental evaluation." Physics in Medicine and Biology **55**: 1453-1473.

Tong, S., A. M. Alessio, et al. (2009). Evaluation of noise properties in PSF-based PET image reconstruction. Nuclear Science Symposium Conference Record (NSS/MIC), IEEE.

Townsend, D. (2008). "Multimodality imaging of structure and function." Physics in medicine and biology **53**(4): R1.

Townsend, D., T. Beyer, et al. (2003). PET/CT scanners: a hardware approach to image fusion, Elsevier.

Townsend, D., T. Beyer, et al. (1998). The SMART scanner: a combined PET/CT tomograph for clinical oncology, IEEE.

Townsend, D. W. (2008). "Dual-modality imaging: combining anatomy and function." Journal of Nuclear Medicine **49**(6): 938-955.

Valk, P. E. (2003). Positron emission tomography : basic science and clinical practice. London; New York, Springer.

Walker, M. D., P. J. Julyan, et al. (2009). Bias in iterative reconstruction of low-statistics PET data: benefits of a resolution model, IEEE.

Warburg, O., K. Posener, et al. (1931). "The metabolism of the carcinoma cell." The Mechanism of Tumors: 129–169.

Watson, C. (2000). "New, faster, image-based scatter correction for 3D PET." Nuclear Science, IEEE Transactions on **47**(4): 1587-1594.

Watson, C., D. Newport, et al. (1996). "A single scatter simulation technique for scatter correction in 3D PET." Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine **4**: 255–268.

Workman, R. and R. Coleman (2006). PET/CT essentials for clinical practice, Springer.

Yao, J. and H. Barrett (1992). Predicting human performance by a channelized Hotelling observer model. International Society for Optics and Photonics, San Diego.

Vita

Joshua Schaefferkoetter was born near Kansas City, MO, to the parents of Donald and Alice Schaefferkoetter. He is the first of three children, with one brother Noah and sister Anna. In 1987, he and his family moved to Knoxville, TN, where he attended Fountain City Elementary School, Gresham Middle School, and then graduated from Central High School. After graduation, he went to the University of Tennessee to pursue an undergraduate degree in Physics; he graduated in 2005, and it was in this year that he was introduced to the Molecular Imaging and Translational Research (MITR) program at the UT Medical Center. The following year, he enrolled in graduate school at the University of Tennessee, and in 2008, he officially joined the MITR program, with David Townsend as his mentor.