8-2007

Validation Study of ReFace (Reality Enhanced Facial Approximation by Computational Estimation)

Diana Kim Moyers
University of Tennessee - Knoxville

To the Graduate Council:

I am submitting herewith a thesis written by Diana Kim Moyers entitled "Validation Study of ReFace (Reality Enhanced Facial Approximation by Computational Estimation)." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Arts, with a major in Anthropology.

Richard L. Jantz, Major Professor

We have read this thesis and recommend its acceptance:

Murray K. Marks, Janice Harper

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

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

To the Graduate Council:

I am submitting herewith a thesis written by Diana Kim Moyers entitled "Validation Study of ReFace (Reality Enhanced Facial Approximation by Computational Estimation)." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Arts with a major in Anthropology.

Richard L. Jantz
Major Professor

We have read this thesis
and recommend its acceptance:

Murray K. Marks

Janice Harper

Accepted for the Council:
Carolyn Hodges
Vice Provost and
Dean of the Graduate School

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

Validation Study of ReFace
(Reality Enhanced Facial Approximation by Computational Estimation)

A Thesis
Presented for the
Master of Arts
Degree
The University of Tennessee, Knoxville

Diana Kim Moyers
August 2007

Dedication

To Dr. Bill Bass, who inspired me and countless others to follow his path.

Acknowledgments

Many people provided assistance, support, and guidance during this research project. The completion of this thesis would not have been possible without them. I express my appreciation and thanks to the following:

This research was supported in part through the FBI's Visiting Scientist Program, an educational opportunity administered by the Oak Ridge Institute of Science and Education (ORISE). This research was also supported in part by a grant from the William M. Bass Endowment.

Funding for the CT scans was arranged by Dr. Mohamed R. Mahfouz, Center for Musculoskeletal Research, Center Technical Director and Director, Computational Bioengineering, Joint University of Tennessee/Oak Ridge National Laboratory. Coordination and scheduling for the scans was done by Megan Moore, use of the CT machine was donated by Dr. Kent Hudson, and the skull scans were done by Todd Malone. Assistance in packaging and transport was done by Donna McCarthy.

The members of my thesis committee: Dr. Janice Harper, for insightful comments and a cultural perspective; Dr. Murray Marks, for providing opportunities for learning beyond the classroom; and Dr. Richard Jantz, for supporting my request to conduct research in a nontraditional area.

From the Federal Bureau of Investigation: Mentors, Dr. Heather Peters (Phase 1) and Phil Williams (Phase 2). Myke Taister, for his continued support and for the artistic renderings of hair and eyes on the computerized facial approximations. The 103 volunteers who participated in the validation testing.

K. Karl Spalvins, whose advice, guidance, and support helped me achieve my academic goals.

Abstract

ReFace (*Reality Enhancement Facial Approximation by Computational Estimation*) is a prototype facial approximation software program developed by the Federal Bureau of Investigation (FBI) in conjunction with GE Global Research. The prototype extrapolates an "approximation" of a face from a skull using a database of computed tomography (CT) scans of living individuals. The test set consisted of CT scans of 53 articulated human skulls from the William M. Bass Donated Skeletal Collection and the William M. Bass Forensic Skeletal Collection, which are curated at the University of Tennessee in Knoxville.

Through the Federal Bureau of Investigation's Visiting Scientist Program, an educational opportunity administered by the Oak Ridge Institute of Science and Education (ORISE), the researcher conducted an independent validation of this software in two phases. *Phase 1* tested and evaluated the software performance, resulting in improvements to the software and the development of standardized protocol for articulation, packaging, and preparation of human skulls for CT scans. *Phase 2* validated the accuracy of the software in the production of facial approximations from human skulls using face pools and resemblance ratings.

In *Phase 2*, computerized facial approximations were visually compared with antemortem photographs by four participant groups (N = 103). Ten test subjects of European ancestry (six females and four males) were selected for a photographic validation by face pool and resemblance rating validation tests. Participants were asked to choose the face pool photograph that most closely resembled the facial approximation produced by ReFace. In the second test, the same volunteers were asked to rate (on a scale of 1 to 5) how closely ReFace facial approximations of target subjects resembled an antemortem photograph.

In the Face Pool Validation Test, nine out of ten target subjects were correctly identified above random chance, and the frequency distribution was statistically above chance expectations for nine out of ten target subjects ($p < .01$). The mean hit rate for all subjects was 24% (10% above random chance). There were no significant differences in the hit rates between male participants (67%) and females participants (33%), or between participant groups. All participants were non-experts. Male target subjects received higher numbers of correct responses than female target subjects. The overall ratings for the Resemblance Rating Validation Test were 13% none, 24% slight, 22% approximate, 25% close, and 16% strong. The majority of subjects were rated as close resemblance (six subjects), strong resemblance (one subject), approximate resemblance (one subject), and slight resemblance (one subject). The foil comparison received an equal number of ratings for no resemblance (30.5%) and slight resemblance (30.5%).

Preface

In July of 2000, a watershed event in the area of craniofacial identification occurred when leading researchers from around the world gathered in Washington, DC, to attend a conference hosted by the Federal Bureau of Investigation. Research results from the 9th Biennial Scientific Meeting of the International Association for Craniofacial Identification (IACI) have been gathered and published in *Computer-Graphic Facial Reconstruction,* edited by John G. Clement and Murray K. Marks (2005). Preliminary research for ReFace, the subject of this research project, can be found in Chapter 8: Face Reconstructions Using Flesh Deformation Modes.

Preliminary research and ideas presented at this conference have continued to develop, and some have become realities. Both the quality and quantity of craniofacial tissue depth knowledge have been increased by use of non-invasive techniques. New tissue depth measurements have been collected on living individuals from diverse geographic and ancestral groups. Databases of computed tomography (CT) scans of living individuals have been and continue to be collected. The importance of facial perception and recognition in the future of facial approximation research has been recognized.

With the joining of computer and medical technologies, forensic facial approximation is no longer the purview of forensic artists and forensic anthropologists. Progress in medical and dental technologies, computer graphics, facial perception and recognition, information technology, medical engineering, biotechnology, and computer science have all contributed to the continued research and development of new and better ways of identifying unknown victims.

Table of Contents

Chapter 1: **Introduction** . . . 1
Research Problem . . . 1
Theoretical Orientation . . . 2
Project Beneficiaries . . . 2
Research Goals . . . 2

Chapter 2: **Review of Existing Literature** . . . 4
Background and Terminology . . . 4
Historical Review of Facial Approximation . . . 4
Traditional Methods of Facial Approximation . . . 5
Two-Dimensional Methods . . . 5
Three-Dimensional Methods . . . 5
Unsolved Problems with Facial Approximations . . . 8

Chapter 3: **Introduction of Computer Methods** . . . 16
Computerized Facial Approximation Prototypes . . . 16
Facial Perception and Recognition . . . 26
Future Development . . . 28

Chapter 4: **Overview of Prototype (ReFace) Validation** . . . 29
Development and Validation History . . . 29
ReFace Database and Research Test Set . . . 29
Software Operation . . . 32

Chapter 5: ***Phase 1*: ReFace Prototype Program Testing** . . . 35
Overview of Prototype Software Testing (*Phase 1*) . . . 35
Methods and Materials: *Phase 1* . . . 35
Articulation, Packaging, and CT Scanning of UTK Test Skulls . . . 36
Results of *Phase 1* Testing . . . 37

Chapter 6: ***Phase 2*: ReFace Prototype Preliminary Validation** . . . 39
Introduction . . . 39
Methods and Materials . . . 39
Target Subjects . . . 39
Results of Preliminary *Phase 2* Testing . . . 42

Chapter 7: **Development of Research Design for *Phase 2* Validation** . . . 46
Introduction . . . 46
Summary of Key Research Articles . . . 46
Selection of Photographs for Comparison Face Pool . . . 48
Face Pool Design . . . 49
Resemblance Rating Design . . . 51
Design of Validation Testing Kits . . . 51

Chapter 8: ***Phase 2*: Final ReFace Prototype Validation** 55
Introduction 55
Methods and Materials 55
Selection of Target Subjects for Validation 55
ReFace Prototype Beta Version (6.0) 56
Face Pool Groups 58
Participant Groups 6 0
Administration of Validation Tests 6 0

Chapter 9: ***Phase 2*: Results of Final ReFace Prototype Validation** 62
Validation Test 1: Face Pool 62
Face Pool Groupings 66
Comparison of Participant Groups 69
Summary of Face Pool Results 71
Validation Test 2: Resemblance Rating 72
Resemblance Ratings by Face Pool Groupings 72
Comparison of Participant Groups 78
Summary of Face Pool and Resemblance Rating Results 80
Discussion 80
Conclusions 84

List of References 87

Appendixes
Appendix A: *Phase 2* Validation Testing Materials 98
A-1 Handout of general instructions for participants in *Phase 2* photographic validation. 99
A-2 Participant instructions for Validation ONE: Face Pool 100
A-3 Face-Pool participant answer sheet 101
A-4 Participant instructions for Validation TWO: Resemblance Rating . 103
A-5 Resemblance Rating participant answer sheet 104

Appendix B: Demographics of Test Subjects and ReFace Database 106
B-6 University of Tennessee Test Subject Demographics: All populations (N = 53) 107
B-7 University of Tennessee Test Subjects Demographics: American Females (European Ancestry) N = 19 108
B-8 University of Tennessee Test Subjects Demographics: American Males (European Ancestry) N = 30 109
B-9 ReFace Database Statistical Data: American Males (European Ancestry) N = 49 110
B-10 ReFace Database Statistical Data: American Females (European Ancestry) N = 49 111

Appendix C: *Phase 1*: ReFace Beta Version Software Validations 112
ReFace Beta Version 1.0 113
Validation I (Beta Version 2.0) 113
Validation II (Beta Version 3.0) 114

Validation III (Beta Version 3.1) 116
Validation IV (Beta Version 4.0) 120
Validation V (Beta Version 5.0) 120
Validation V (Beta Version 5.1/5.2) 124

Appendix D: Protocol Developed During Software Evaluation and Testing . . . 130
Articulation of Mandible and Cranium 131
Testing of Articulation and Packing Materials 132
Protocol: Articulation 134
Protocol: Packaging 134
Protocol: Skull CT Scans 135
Protocol: Removal of Dental Noise 135

Vita 137

List of Tables

Table		Page
4.1	University of Tennessee Validation Subjects: Ancestry and Age Distribution	30
4.2	ReFace Database: Ancestry and Age Distribution	33
7.1	Validation Resemblance Rating Scale	52
8.1	Demographics of Subjects in Face Pool Validation	59
9.1	Face Pool Validation Summary of Chi Square Values	65
9.2	Face Pool Validation: Binomial Probability Calculation Summary	67
9.3	Comparison of Binomial Probability and Actual Results (%)	67
9.4	Resemblance Rating Validation Test (Modes and Medians)	73
9.5	Frequency of Resemblance Ratings by Subject (N = 101)	73
9.6	Frequency Distribution with Grouped Categories	76
C.1	Summary of Reconstruction Tests (Beta 5.1/5.2)	127

List of Figures

Figure		Page
2.1	Articulated Skull with Traditional Tissue Depth Markers Applied	6
2.2	Example of American 2D Method of Facial Approximation	7
2.3	Examples of American 3D Method of Facial Approximation	9
2.4	Examples of Russian Method of Facial Approximation	10
2.5	Example of Combination Method of Facial Approximation: Traditional Clay	11
2.6	Example of Combination Method of Facial Approximation: Virtual Clay	12
2.7	Examples of Medical Imaging	14
3.1	Example of Facial Approximation by Evenhouse et al 1992	17
3.2	Example of Facial Approximation (traditional markers) by Evison 1996	18
3.3	Example of Facial Approximation by Evison 1996	19
3.4	Example of Facial Approximation by Quatrehomme et al 1997	21
3.5	Example of Facial Approximation by Nelson and Michael 1998	22
3.6	Example of Facial Approximation by Vanezis et al 2000	23
3.7	Example of Facial Approximation by Marks et al 2000	24
3.8	Example of Facial Approximation by Jones 2001	25
3.9	Example of Facial Approximation by Claes et al 2006	27
4.1	University of Tennessee Phase 1 Testing Sample by Age, Sex, and Ancestry	31
4.2	ReFace Database by Age, Sex and Ancestry.	34
6.1	Facial Approximations by ReFace with Weight Range of Thin, Average, and Heavy for Subject 37-02D	40
6.2	ReFace Approximations: Dental Artifacts Image Distortion (before) and DICOM Corrected (after)	41

6.3 Evolution of ReFace Prototype: Bone and Skin Models from Beta Versions 1.0, 3.1, and 5.1. 43

6.4 Comparison of ReFace Approximations with Antemortem Photographs and with Artistic Addition of Hair and Opened Eyes to Computer Images 44

7.1 Design Layout for Photographic Face Pool Validation 50

7.2 Design Layout for Photographic Resemblance Rating Validation 53

9.1 Face Pool Results for All Participant Groups (N = 103) 63

9.2 SPSS Output for Chi Square Frequencies for Ten Target Subjects. 64

9.3 Face Pool Validation Results by Participant Groups (N = 103) 70

9.4 Resemblance Rating Responses by Category for All Participant Groups: Number of Responses (N = 101). 74

9.5 Resemblance Rating Responses by Category for All Participant Groups: Responses by Percent (N = 101). 75

9.6 Resemblance Ratings by Percent by Participant Groups 79

Glossary

(As used in this project)

Admin: Administrator. Programming menus of the ReFace program are available only to the administrator, and not to the operator of the program. Testing of the prototype included verification that the administrator (Admin) functions were working properly.

Antemortem Photographs: Photographs of the target subject taken before death (as opposed to photographs taken after death, such as in a morgue or crime scene).

ARF: The Anthropological Research Facility at the University of Tennessee in Knoxville, Tennessee, founded by William M. Bass.

Assessors: Participants in the evaluation that were asked to make choices in the comparison of computerized facial approximations with antemortem photographs, (also known in different studies as judges, raters, or volunteers).

Bone.vtk: The file name for a three-dimensional skull model generated by ReFace after compiling CT DICOM slices from an unknown skull.

Current Estimate: ReFace viewing window of a completed facial approximation.

Decomposition: A process that begins immediately after death and continues in stages over time until full skeletonization.

Defleshed: Removal of the soft tissue from the skull.

Dental Noise: Flashes of light in CT slice images from metal dental fillings, or metal bases on dentures, crowns, screws, or other dental work. These flashes of light are interpreted as bone by ReFace and cause projections to extend from the mouth area. The extensions, or spikes, cause distortions of the skin overlay.

DICOM: Digital Imaging and Communications in Medicine Standard for distributing and viewing any kind of medical image regardless of the origin. A series of two-dimensional data obtained from computed tomography (CT) scans. ReFace forms a three-dimensional model from the DICOM data called the bone model or bone.vtk.

Dry Skull: A skull that has been defleshed and allowed to dry, or one that has been exposed to environmental elements and is dry to the touch. Skeletal remains are usually dry; decomposing remains are usually wet.

Effect 1 and Effect 2: ReFace slider bars to increase or decrease weight in Beta Versions 1.0 through 5.2. Adjustments can be made in Effect 1 and Effect 2 controls from -10 to +10, which reflect two standard deviations from the average reconstruction.

Extrema: The maximum and minimum of a mathematical function producing a facial characteristic such as gauntness to obesity.

Face Pool: An array of photographs that contain the target individual which is used for comparison with a facial approximation.

Familiar Method: Assessors know the identity of the missing person and look at a face pool to see if the person they know is there. (Usual and customary in forensic identification of missing persons, but difficult to replicate in a scientific setting.)

Foil Comparison: In resemblance tests, a facial approximation is compared with an unrelated photograph (not the target).

Forced Choice Method: Assessors are forced to choose a photograph from a photo pool even if they believe the target subject is not present. Assessors are requested to choose the photograph that most closely resembles the facial approximation, thus it may be implied, but not stated, that the target subject is not there.

IACI: The International Association for Craniofacial Identification.

NCIC: The Federal Bureau of Investigation's National Crime Information Center.

Open Ended Choice Method: Assessors are given the option of selecting "None of the Above" in a photo face pool.

ReFace: Reality Enhancement Facial Approximation by Computational Estimation.

Resemblance Rating: A comparison test to access similarity between a facial approximation and a target (known) individual.

Skin.vtk: The file name of a three-dimensional skin surface overlay on a skull model. The skin model is the facial approximation image generated by ReFace.

Scout Images: DICOM files that usually appear as the first or second in a series of files representing the individual slices from a CT scan. Scout files are not slices, but images of front or profile views of a complete skull.

Target Individual: The actual person represented by a facial approximation in a photo pool. The correct answer to the face pool test is the target (person approximated by ReFace for comparison).

Three-Dimensional (3D): Existing in three dimensions and giving the illusion of depth. 3D facial approximations can be viewed or manipulated in actual (traditional) or virtual (computer) space to see all surfaces of an image.

TMJ: The temporomandibular joint; the articulation point between the mandible and the cranium

Two-Dimensional (2D): Existing in two dimensions and lacking the illusion of depth. 2D facial approximations created by artistic sketching over photographs of a skull with attached craniofacial landmarks.

Unfamiliar Method: Assessors do not know or have never seen any of the people in the photographs used in the experiment. (Used primarily in scientific validation research; not usual or customary in forensic cases.)

UTK: The University of Tennessee at Knoxville, Tennessee.

Wet Skull: Living bone (skulls from the ReFace database from living individuals) or decomposing remains with soft tissue still attached and with fluids still contained in the bone.

Chapter 1
Introduction

Traditional forensic facial approximation is an expensive, time-consuming method of last resort used to help identify unknown skeletal remains. It usually involves building an approximation of a face on a skull, with the goal of recognition by someone who knew the victim. Combining forensic anthropology biological profiles with artistic talent, it is often described as a blending of science and art (Phillips and Smuts 1996, Phillips 2000; Taylor 2001, Tyrrell et al 1997, Wilkinson 2002), although some disagree that any "science" is involved (Stephan 2002a, 2002b, 2002c, 2003a, 2003b; Stephan and Henneberg 2001, 2006; Stephan et al 2003).

Facial approximation can be used for historical, archaeological, and forensic cases. Labor intensive, expensive, and subjective, few of the thousands of human remains found will ever have a facial approximation done. The focus of this research was on forensic approximations, and the technological advances that have led to state-of-the-art prototypes for this application.

Research Problem

Current methods and materials for the identification of unknown skeletal remains by facial approximation are not standardized, cost effective, or objective. In such cases, a forensic anthropologist provides a biological profile (age, sex, ancestry, height, weight), which determines the craniofacial tissue depth "averages" to be used. An artist must then choose from a variety of tissue depth information tables which are inconsistent with each other. Depending on the tissue depth data selected and the skill level of the forensic artist, reconstructions made by different artists on the same skull may vary (Haglund and Reay 1991). Prior research on measurements taken of tissue depths between the bone and skin surface has generally concluded that tissue depths are age, sex, and population specific.

Although forensic anthropologists have collected soft tissue depths since the late 1800s, they have yet to develop or agree upon a standardized set of tissue depth markers, reach a consensus on which craniofacial landmarks should be used, or even their exact locations (Brown et al 2004). The averages of tissue depth can vary depending on the ancestry, sex, and body type. Different populations have different "averages" yet no one person in any given population would necessarily be an "average" specimen of that population.

Studies to determine the thickness of soft tissue were done on cadavers until the middle 1980s (His 1895, Kollman and Büchly 1898, Suzuki 1948, Rhine and Campbell 1980). A forensic artist uses the biological profile and average tissue depths obtained from a chart to sculpt or draw the basic face and artistic "rules-of-thumb" to add soft tissue features (ears, nose, mouth). The result is an "approximation" of how the person looked during life, but because there are no standardized forensic art techniques or anthropological tissue depth tables, the results are subjective and often inconsistent.

Theoretical Orientation

Two diametrically opposing anthropological viewpoints are involved in current forensic facial approximations (scientific versus unscientific). The viewpoint, methodology, and theory of many physical anthropologists is the scientific method, whereby a testable hypothesis is formulated, rigorously tested, and ends with empirically grounded results. This would equate to the rigidly scientific archaeological theory of processualism, advanced by Louis R. Binford, which became known as the New Archaeology of the 1960s. In contrast, forensic artists use unscientific talent, skill, and artistic "rules-of-thumb" to determine the appearance of "imaginary" soft tissue features. This equates with the archaeological theory of post-processualism, which focuses on individuals, cognition, and a humanistic approach. However, forensic facial approximation needs both science *and* art to be successful. This blending equates with the processual-plus theory of Michelle Hegmon, which uses scientific methods for testing those parts of research that can be tested, yet incorporates and also attempts to validate parts of the research which cannot be empirically tested.

Project Beneficiaries

No one should be buried without a name. The primary beneficiaries of such a validated system are victims and their families. Thousands of John and Jane Does are currently housed in university research labs and morgues. Without a name, these unidentified remains will never be identified and reunited with their families. Identification of unknown remains provides closure to families and assists in the identification and prosecution of suspects. Additional beneficiaries include forensic anthropologists, forensic artists, law enforcement, criminal prosecution, human rights organizations, and the criminal justice system.

Traditional methods of facial approximations are labor intensive, and therefore expensive. Limited funding may limit or negate the hiring of a professional forensic artist. A traditional three-dimensional facial approximation can take weeks to complete; computerized facial approximations could expedite this process to a few hours. This reduction in labor would result in substantial time and cost savings over current methods, making the process available to all cases. The ability to provide a facial approximation to the media soon after the discovery of human remains may increase the probability of identification. A rapid, cost-effective method could be utilized at the beginning of an investigation (rather than the last resort at the end of the investigation). Quick identification of the victim by law enforcement is the key to identifying potential suspects. In many jurisdictions, prosecution against a suspect cannot commence until the victim has been positively identified.

Research Goals

A computerized facial approximation system could provide a standardized set of landmarks and tissue depths for use in traditional three-dimensional forensic facial approximations. Additionally, the rules-of-thumb currently used by forensic artists could be objectively tested for accuracy.

The two goals of this research were to test and evaluate the prototype software, ReFace, and then to validate the accuracy of the software by comparison of facial approximations by ReFace with antemortem photographs of target subjects. Both goals were accomplished using CT scans of 53 human skulls of known age, sex, and ancestry from the William M. Bass Donated and Forensic Skeletal Collections. The skulls were CT scanned and entered into ReFace as “unknown skulls” along with the biological profile. The prototype used this information to produce “approximations” for each sample skull. The approximations were then compared to antemortem photographs both for testing and validation purposes.

Phase 1 Software Evaluation: Numerous problem areas were identified during the testing, which were submitted to the program developers for correction and further development. Testing details are provided in Appendix C. Most have been successfully resolved, and additional corrections and enhancements were included in Beta Version 7.0, released shortly after this project was completed. Protocols were developed for the articulation, packaging, and preparation of human skulls for CT scanning (Appendix D).

Phase 2 Validation: Validation of the accuracy of ReFace was completed by the comparison of photographic images to facial images produced from the prototype software. A biometric comparison of the images was planned to quantitatively analyze the degree of similarity between the approximated face and the actual photograph of the living person. However, the facial recognition software needed for this comparison was still in development and was not available for this research.

Facial approximations for ten target subjects were prepared with ReFace Beta Version 6.0. In the first validation test, participants attempted to choose the antemortem photograph represented by the ReFace facial approximation in a Face Pool Validation Test. In the second test, the same participants were then asked to rate the resemblance of the approximation to the antemortem in a Resemblance Rating Test. Results of the testing are provided in Chapter 9.

ReFace provides an automated, computerized, method of building a three-dimensional face from a human skull. Although artistic skills and knowledge were crucial in the design and development of ReFace, it is not anticipated that an operator of the final product will require artistic or anthropological skills. Such a system could provide facial approximations not currently addressed by traditional methods, such as advanced decomposition and situations that result in high volumes of unidentified remains. With advanced decomposition, traditional methods require the removal of the head from the body and defleshing of the skull. The removal of the head from the body may not be legally permissible or culturally acceptable. Mass disasters may leave hundreds of decomposing bodies; genocides may leave hundreds of commingled skeletal remains. A rapid cost effective method for high volume is clearly needed in such circumstances.

Chapter 2
Review of Existing Literature

Despite attempts to standardize traditional techniques, facial approximation has and will always remain on the threshold between art and science. Current research in forensic facial approximation has been greatly assisted by advances in computer science and medical imaging, such as computer tomography (CT), which was first known as CAT (Computed Axial Tomography). CT scans of living persons reveal an exact margin between bone and soft tissue, enabling accurate measurements of tissue depths. Artistic "rules-of-thumb" as to the shape and size of soft tissue features (ears, nose, lips) can be quantified and statistically correlated.

New computerized facial approximation prototypes have the potential to end the art versus science debates in the area of craniofacial identification of unknown human remains. Standardization of tissue depth and anthropometric measurements could result in economical, objective, and fast facial approximations. A brief review of existing methods is provided, as well as an analysis of current problems, current research, potential solutions, and leading prototype development.

Background and Terminology

A common misconception among laypersons, scientists, and the media is that the approximation will be an exact likeness instantly recognizable by anyone when compared to a photograph of the actual subject. However, it is not intended to be and rarely is a precise likeness of the victim (Manhein 1999, Phillips 2000, Tyrrell et al 1997, Ubelaker and O'Donnell 1992). Forensic facial approximations are not intended to be recognized by strangers but rather to spark recognition in someone who *knew the victim* (Ubelaker and Scammell 1992). Rarely will the approximation produce an immediate "hit" as in the movies.

The terms ***facial approximation*** (Craig 2003, George 1987, Haglund and Reay 1991, Reichs and Craig 1998, Stephan 2002a, 2002b, 2002c, 2003b, 2006); ***facial reconstruction*** (George 1987, Evenhouse et al 1992, George 1993, Helmer et al 1993, Hering 2003, Miyasaka et al 1995, Nelson and Michael 1998, Phillips and Smuts 1996, Quatrehomme and Cotin 1997, Shahrom et al 1996, Snow et al 1970, Stoney and Koelmeyer 1999, Tyrrell et al 1997, Vanezis et al 1989, Vanezis et al 2000, Wilkinson 2004b, Wilkinson and Whittaker 2002); ***facial reproduction*** (Rhine 1990, Ubelaker and O'Donnell 1992); and ***facial sculpture*** (Gatliff 1984); have been used interchangeably in the literature to describe the 2D or 3D methods for historical, archaeological, and forensic construction of a face based on a skull.

Historical Review of Facial Approximation

Three-dimensional facial approximation has been traced to 7 B.C. (Pre-Pottery Neolithic) when death masks were found in Jericho in the Jordan Valley (Prag and

Neave 1997). In the 1800s, attempts at approximation were made primarily for historical faces. The first approximation based on scientific evidence of soft tissue thickness is often attributed to the German anatomist, Wilhelm His, who constructed a face in 1895 on the skull of Johann Sebastian Bach (Gruner 1993, His 1895, Prag and Neave 1997). In 1898, Kollman and Büchly created a table of average tissue thicknesses of male and female European Caucasians based upon scientific methods. This was the only data available until publication in 1962 of *The Human Skeleton in Forensic Medicine* (Joyce and Stover 1991, Krogman 1962).

Traditional Methods of Facial Approximation

Two Dimensional Methods

Artistic Renderings

The two-dimensional method is synonymous with the work of Karen Taylor, which is defined and illustrated in her reference manual, *Forensic Art and Illustration* (2001). As with three-dimensional reconstructions, a clean skull is articulated, and tissue depth markers attached to the skull surface (see Figure 2.1). The skull is then aligned and photographed in front and profile views, and these two-dimensional photographs are attached to drawing boards. The facial approximation is created by sketching around the tissue depth markers and then adding soft tissue features and hair to present a lifelike appearance. Figure 2.2 is an example of one of the many cases from her book. This method is ideally suited for very fragile skulls that may be unable to support the weight of attaching clay, and also provides a method of creating a facial sketch of recently deceased individuals based on photographs. However, extensive decomposition of the face would still require defleshing of the skull, as would three-dimensional methods.

Skull Photo Superimposition

A method of forensic identification involving the skull, but not the forensic artist, which may or may not involve a computer or a video camera, is facial superimposition. The position of a skull and a photograph are adjusted with the image reflecting off mirrors. The result is the merging of the skull and the photograph into one image. (Joyce and Stover 1991, Zonderman 1999, Taylor and Brown 1993, Byers 2002). In a case without an available photograph for comparison, a positive identification was made by using photographic still from home video movies superimposed over images of dental remains (Marks et al 1997).

Three Dimensional Methods

The three current methods are the American (Tissue Depth) Method, generally associated with the work of Betty Pat Gatliff (Joyce and Stover 1991, Schwartz 1993, Taylor 2001); the Russian Anatomical Method, developed in the 1930s by Russian anatomist and forensic anthropologist Gerasimov (Gerasimov 1971, Taylor 2001, Wingate 1992, Zonderman 1999); and the Combination Method, popularized by Richard Neave in Europe (Prag and Neave 1997, Taylor 2001). Facial approximations have ranged from a historical clay bust of Francisco Pizarro to the nine unidentified victims of serial killer John Wayne Gacy (Joyce et al 1991, Maples 1994).

Figure 2.1: Articulated skull with traditional tissue depth markers applied (skull from the University of Tennessee William M. Bass Donated Skeletal Collection).

American 2D Method

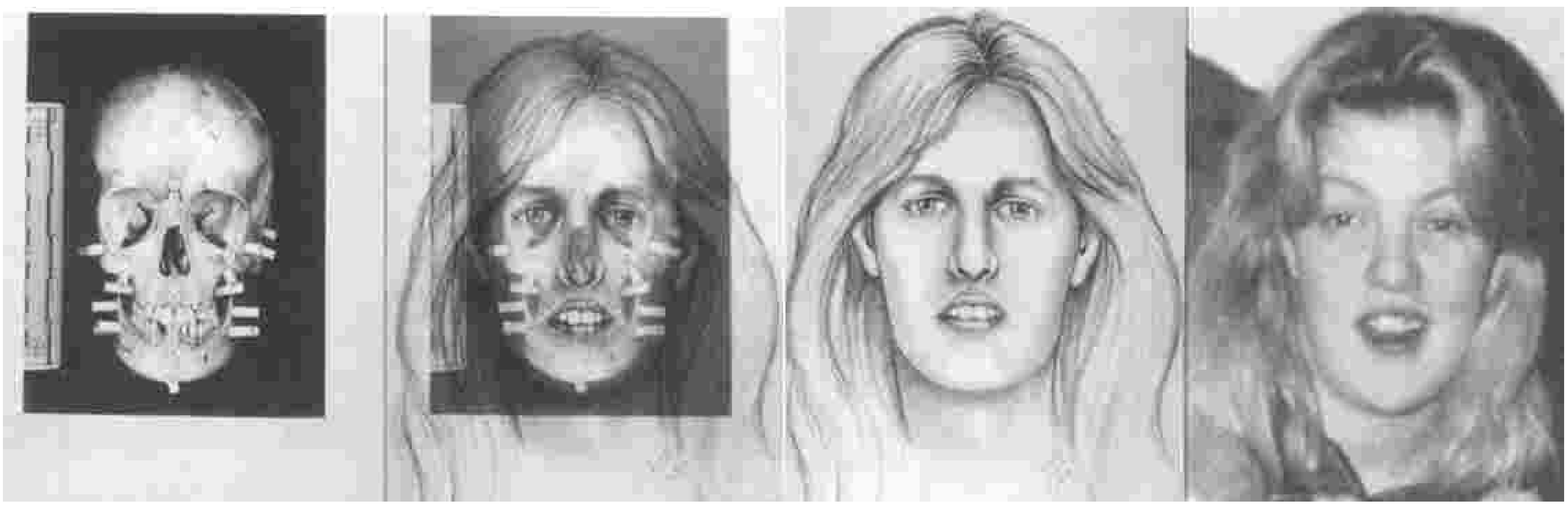


Figure 2.2: Example of American 2D Method of Facial Approximation by Karen Taylor (modified from Taylor 2001, page 385).

Tissue Depth Method (American)
The Tissue Depth Method (also known as the American Tissue Depth Method) is generally associated with the work of Betty Pat Gatliff. As with the 2D American method, a series of tissue depth markers are created from a material that can be attached to the skull, such as plasticine or erasers. Each piece represents an average thickness of skin, muscle, and fat for a specific craniometric landmark on the skull. Each marker is then connected to the others around it with strips of material such as clay, plasticine, or wax. Spaces between the strips are filled in and smoothed to approximate the contours of the face. (Schwartz 1993, Taylor 2001). Forensic experts, Dr. Clyde Snow (a forensic anthropologist) and Betty Pat Gatliff (a forensic artist) worked together to create facial approximations that ranged from a historical clay bust of Francisco Pizarro to the nine unidentified victims of serial killer John Wayne Gacy (Joyce 1991, Maples 1994). (See Figure 2.3).

Anatomical Method (Russian)
The Anatomical Method (also known as the Russian Anatomical Method) involves the reconstruction of facial muscles and tissues. This method is primarily associated with Gerasimov, a Russian anatomist and forensic anthropologist working in Moscow. Gerasimov also did historical reconstructions (Ivan the Terrible and Attila the Hun) and wrote a novel about this work, *The Face Finder* (Wingate 1992). The novel, *Gorky Park*, by Martin Cruz Smith was based on Gerasimov's work. Decomposed skulls of three people found in Gorky Park were given to Gerasimov who used insects to remove the remainder of the soft tissue from the skulls (Zonderman 1999). (See Figure 2.4).

Combination (Manchester) Method
The Combination Method combines the Anatomical and the Tissue Depth approximation methods. It was popularized by Richard Neave in his archaeological and historical approximations, including Lindow Man, Philip II, and Yde Girl (Prag and Neave 1997). Facial muscles are built over a plaster or actual skull (as in the Russian Method) and cross-referenced with traditional tissue depth markers (as in the American Method) (Prag and Neave 1997, Taylor and Angel 1998, Wilkinson 2004b). Figure 2.5 illustrates this method using actual clay (from Milner et al 2000), and Figure 2.6 illustrates this method using computer assisted "virtual" clay (Wilkinson et al 2006).

Unsolved Problems with Facial Approximation

Historical and archaeological facial reconstructions have the luxury of time, while forensic cases are rarely blessed with either time or money. Mass disasters (natural and manmade) and human rights violations result in thousands of unidentified human remains. There are simply not enough qualified forensic artists, time, or money to give faces to all of them.

Time, Expense, and Subjectivity

The three primary problems with traditional methods of forensic facial approximation are time, expense, and subjectivity. Traditional approximations are labor intensive (and therefore expensive) and require artistic skill. However, it is the subjectivity that is of

American 3D Method

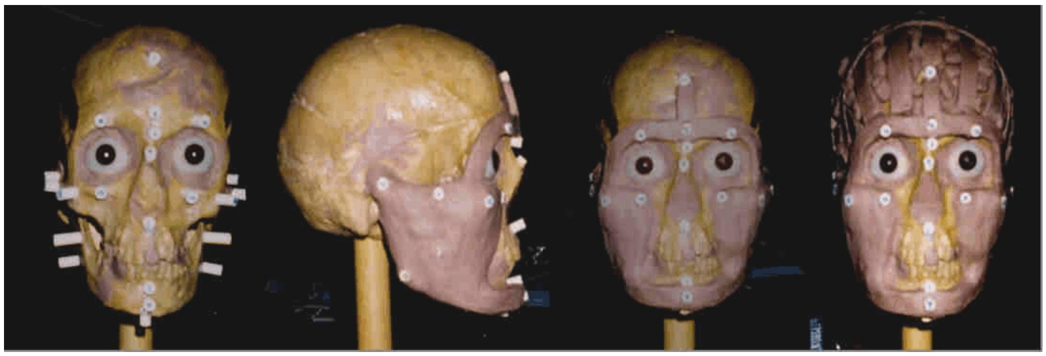


Figure 2.3 (A): American Method of Facial Approximation (3D). Tissue Depth markers are attached to the skull, artificial eyes set in the eye orbits, and clay is used to fill in between the markers. (Models by Diana Moyers).

Figure 2.3 (B): Example of 3D Facial Approximation by Betty Pat Gatliff (in Taylor 2001, page 462).

Russian 3D Method

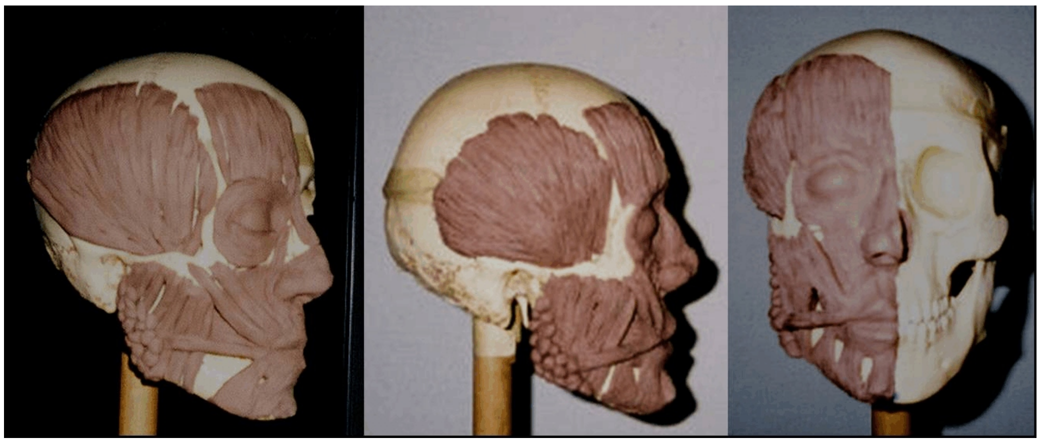


Figure 2.4 (A): Example of Russian Method of Facial Approximation (models done by Diana Moyers).

Figure 2.4 (B): Example by Gerasimov of the Russian Method of Facial Approximation. (From Taylor 2001, page 342).

Combination (Manchester) 3D Method

Figure 2.5: Combination (Manchester) Method of Facial Approximation (3D) done with clay. (Modified from Milner et al 2000).

Combination (Manchester) 3D Method

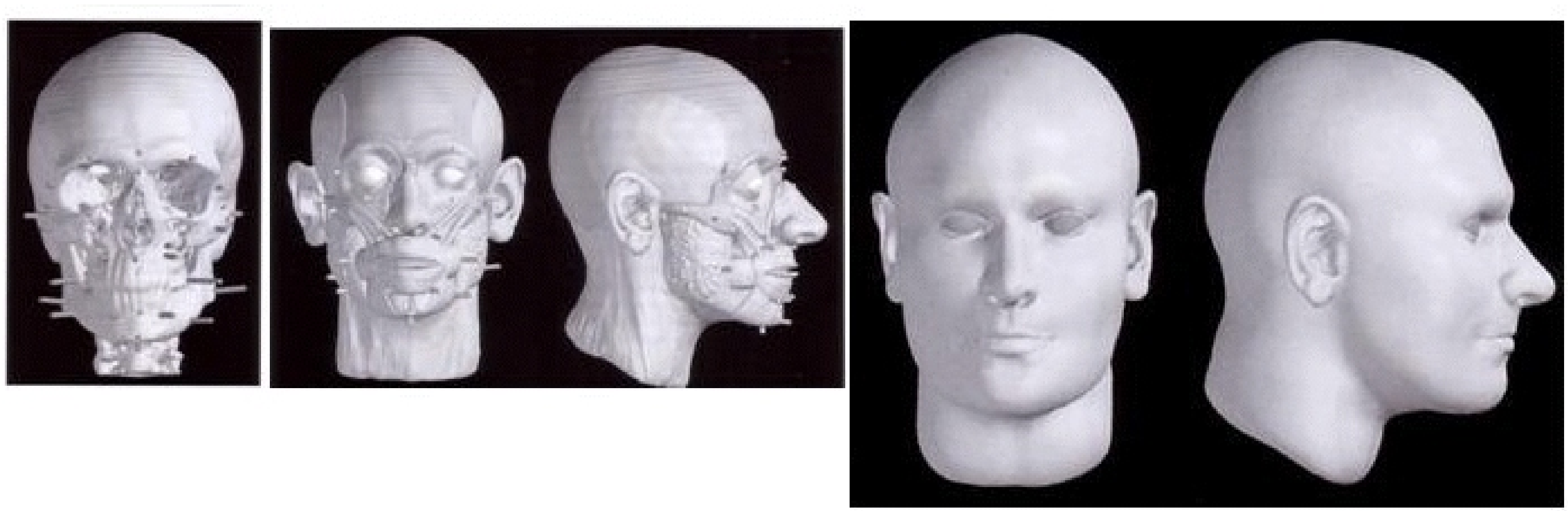


Figure 2.6: A computer assisted facial approximation using the Combination (Manchester) Method of Facial Approximation (3D) done with virtual clay. (Modified from Wilkinson et al 2006).

the most concern. Although age, ancestry, and musculature can be reasonably estimated, other items of importance in identification can not. There are no reliable indicators of weight, hair color or texture, or the shape of soft tissues. Therefore, artists must use artistic "rules-of-thumb" to approximate the missing soft tissue features of eyes, ears, and lips, and intuition for the hair color, hair style, and eye color (Taylor 2001). These artistic methods cannot be scientifically measured, quantified, or tested, and thus are deemed "unscientific" by many anthropologists.

Reliance on Tissue Depth Tables

Tissue depths vary between populations as well as within populations (Krogman 1962, Rhine and Campbell 1980, Aulsebrook et al 1995, Aulsebrook et al 1996, Phillips and Smuts 1996, Rhine 1998, Manhein et al 2000, Taylor 2001, Simpson and Henneberg 2002, Williamson et al 2002, Sahni et al 2002). Which tissue depth chart to use is determined by the individual artist making the facial approximation. Larger sample sizes may not overcome the population specific aspect of tissue depths.

Tissue Depth Tables from Cadavers

Traditional tissue depth tables developed by anthropologists are biased in several ways: use of cadavers for measurements, using an "average" tissue depth, small sample sizes and lack of population diversity. Additionally, there are no standardized tissue depth tables. Traditional data were generally derived from small samples of cadavers with measurements taken from a horizontal position. Decomposition starts immediately after death, and horizontal positioning causes the skin to sag due to gravity. Tissue depth charts developed in the 1980s by Rhine and Campbell attempted to correct for this by measuring unembalmed individuals that had been deceased for 12 or fewer hours (Rhine and Campbell 1980, Vanezis and Vanezis 2000). Embalming was shown to cause significant increases in facial soft-tissue depths (Simpson and Henneberg 2002). Tissue depths taken from embalmed cadavers (minimum 6 months to more than 12 months after embalming) showed minimal differences between males and females (Domaracki and Stephan 2006).

Tissue Depths from Living Subjects

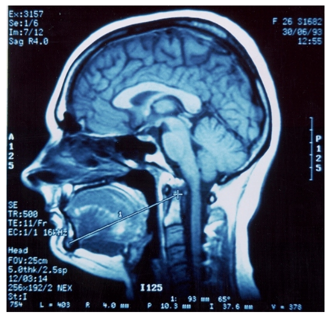
The ability to collect more accurate tissue depth measurements has been aided by the development of medical imaging technology and methods. Methods used in previous research include cephalometric radiographs, magnetic resonance imaging (MRI), ultrasound, and computed tomography (CT) (see Figure 2.7).

Cephalographs and MRI

George (1987) used lateral cephalographs to establish midline soft tissue depths to construct facial approximation profiles on living subjects. Aulsebrook and colleagues (1996) combined cephalographs with ultrasound measurements to develop profiles for African Zulus. Sexual differences were found to be highly significant in an Indian population using MRI scans, and a comparison with different populations supported population specific tissue depths (Sahni et al 2002).

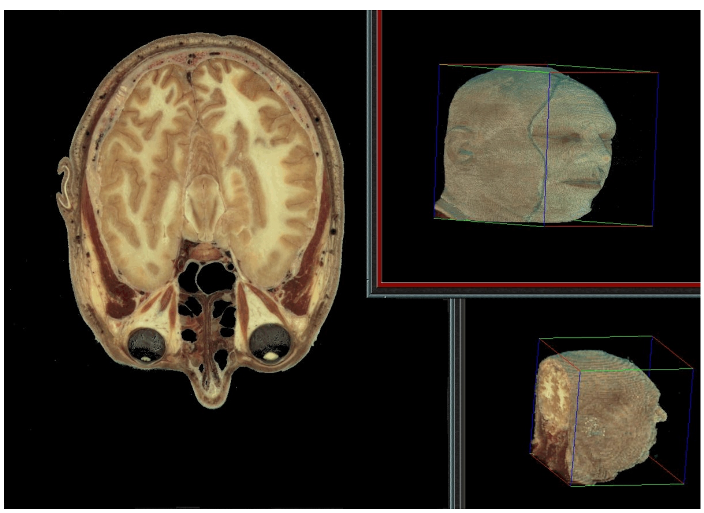
Medical Imaging

Cephalograph of Skull (from Milner et al 2000).



CT Scan of Living Person (from Evison 1996).

MRI Scan of Living Person (from Sahni et al 2002).



MRI of Living Person (from Marks et al 2000)

Figure 2.7: Examples of medical imaging scans.

Ultrasound
Ultrasound can provide *in vivo* tissue depth measurements of craniometric landmarks covered by soft tissue. However, craniometric landmarks cannot be seen through skin and the locations must be estimated. Consistency in replication of landmark locations by different operators could be difficult to achieve. Manhein and colleagues (2000) took ultrasound measurements on American subjects (551 children and 256 adults) at 19 facial locations (13 traditional landmarks and six not previously measured by other researchers). Their findings indicate that some tissue depth points on the face are significant for age, sex, and race, but most tissue depth means were similar among different age groups, races, and sexes.

Computer Tomography (CT)
In the 1990s, forensic facial approximation research methods benefitted from advances in technology originally designed for facial surgeries on living patients. The most promising development was the use of three-dimensional living images from CT scans. As the exact margins between bone and soft tissue could be detected, accurate tissue depth measurements could be made (Tyrrell et al 1997). CT scanners have the ability to collect depths on all the points on a skull, not just specifically defined craniometric markers. Scans provide simultaneous location of cranial landmarks and visualization of the overlaying soft tissue (DeGreef and Willems 2005). Phillips and Smuts (1996) used CT scans to measure facial tissue thickness of a mixed population group in South Africa. Marked differences were found in this population to the tissue depth charts of American Whites and Blacks from Rhine and Campbell (1980).

Accuracy of Methods
Rocha and colleagues (2003) tested linear measurements from bone and soft tissue on CT scans and concluded that the method provided high quality images with a low standard of error for obtained measurements. Kim and colleagues (2005) compared CT scans measurements with physical puncturing of cadavers, and found the CT scans accurately measured soft tissue thickness.

DeGreef and colleagues (2006) compared ultrasound measurements to CT scan measurements and found statistically significant differences in the masseter region, which was attributed to whether the head was horizontal or upright when the measurements were taken. Statistically significant differences were found in this study of Europeans as compared with previous American studies (both cadaver and living). It has been consistently found that tissue depths are population specific; therefore the comparison of American “Caucasians” and European “Caucasians” may be indicative of the variation between these two populations, rather than the accuracy of the different measuring methods.

Chapter 3
Developments and Trends

Introduction of Computers (speed, accuracy, objectivity)

The underlying assumptions for facial approximation are that every human has a unique face, and a facial approximation built over an unidentified skull will provide a reasonable likeness of the living face. The underlying, but as yet unproven, assumption for computerized facial approximations are that computers are more objective than people, and could therefore provide more objective facial approximations than forensic artists.

To address the three primary problems of time, expense, and subjectivity, computers were introduced into the process to create an accurate, flexible, and repeatable system (DeGreef and Willems 2005). The FBI and the Smithsonian Institute were one of the first to combine traditional methods with computers to produce computer-*assisted* facial approximations (Phillips and Smuts 1996, Ubelaker 2000, Ubelaker and O'Donnell 1992). Speed was gained and accuracy improved with this method. However, subjectively was still a problem.

Computerized Facial Approximation Prototypes

Early research in the development of automated computer facial approximation methods initially combined traditional tissue depth measurements with manually digitized skulls. The first known method of facial approximation that merged tissue depths with computers was published in 1989 by Vanezis and colleagues (Vanezis et al 1989, DeGreef and Willems 2005, Turner et al 2006). After a skull was manually digitized, tissue depth markers were digitally added, and computational methods were used to connect the markers.

A computer method developed by Evenhouse and colleagues (1992), attempted to quantify the face by reflected the metric relationships of facial features. A “polygonal mesh” was created by connecting 37 control points of soft tissue and corresponding bony landmarks, which were used to map an average face over an unknown skull. The resulting model was a faceted skin surface that was suggestive of a human face (see Figure 3.1).

Computerizing traditional methods of facial approximation, even if fast and repeatable, were still relying on old data. At the University of Sheffield, a comprehensive project was designed that incorporated 3D laser scans of skulls, updated tissue depth measurements from CT scans, and a computerized facial feature image library. It was anticipated that an animated 3D facial approximation could be viewed over the internet. (Evison 1996, Tyrrell et al 1997, Evison et al 1998, Evison and Green 1999, Green and Evison 1999, Evison 2000). Figure 3.2 is an example of a computerized facial approximation using traditional tissue depth marker placement; Figure 3.3 is an example of a facial approximation completed by the planned new method.

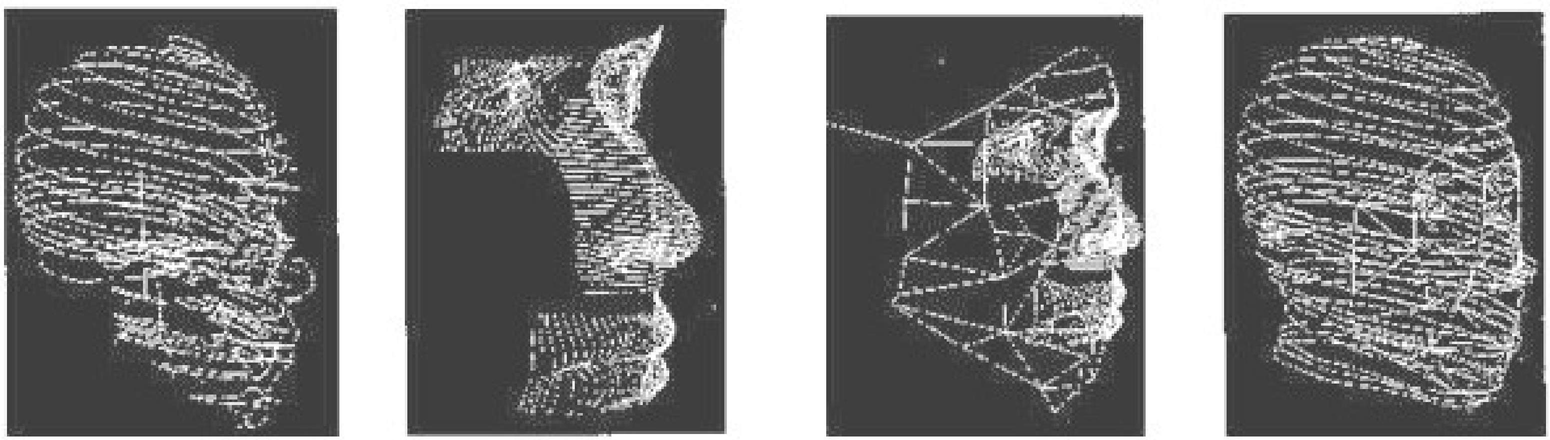


Figure 3.1: Facial approximation methods using metric relationships of facial features. (Modified from Evenhouse et al 1992).

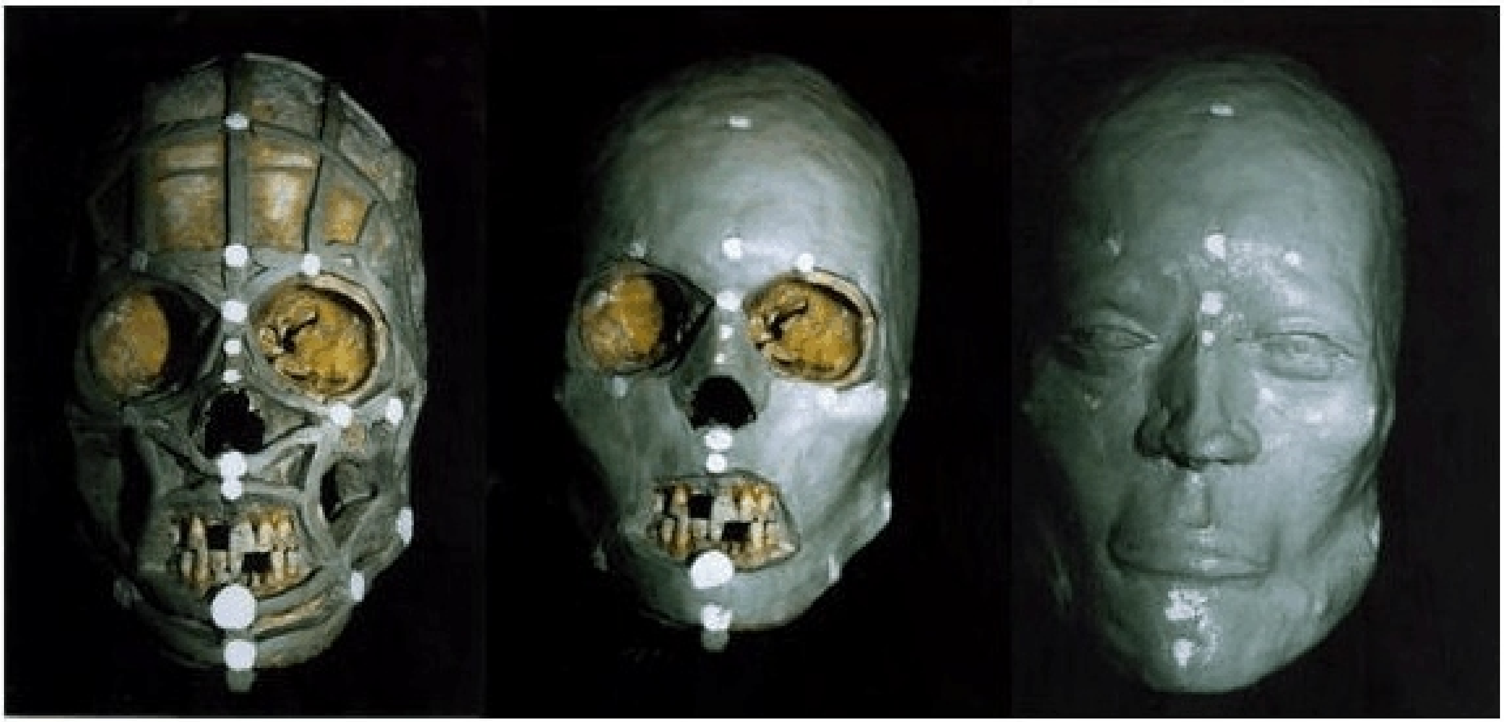
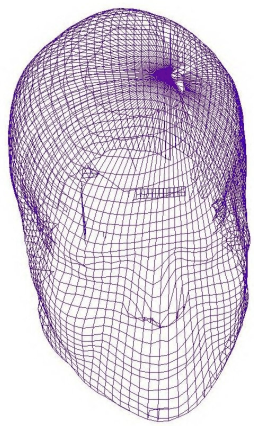
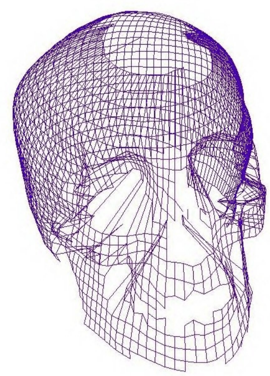


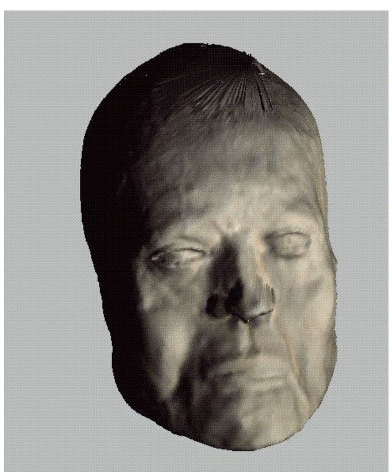
Figure 3.2: Computer facial approximation using traditional tissue depth markers (from Evison 1996).



Wireframe Matrix of
Skull.



Wireframe Matrix of
Face.



Rendered Image of Facial
Reconstruction.

Figure 3.3: University of Sheffield Prototype for Computerized Facial Reconstruction (modified from Evison 1996).

A different method for facial approximation developed by Quatrehomme and colleagues (1997) relied on the principle of deformable models. The assumption was that two skulls with similar forms would have similar faces. Skull and facial models of two cadaver heads were created; one data set was used as a reference and the other as an unknown. This process used a statistical formula as opposed to specific landmarks or average facial tissue thicknesses. The resulting facial approximation is compared with the actual death mask in Figure 3.4.

Volume deformation was also used in a prototype developed by Nelson and Michael (1998), in which CT scans were used to develop a database of reference heads. An unknown skull was compared to reference heads with similar age and sex ranges. Control points between the two heads were established and the reference head was "deformed" over the unknown skull. The facial approximation produced by this prototype is illustrated in Figure 3.5.

With the development and availability of medical imagery, updated tissue depth information became available. Initially this data was manually compiled, but laser scans automated the import of three-dimensional skull images, and methods of facial approximation incorporated additional craniofacial landmarks. The development of databases sought to compare an unknown skull with a known reference base. Laser scanning methods developed by Moss and colleagues (1989) were later incorporated by Shahrom and colleagues (1996) with the addition of superimposition between an "average" face selected from a database and an unknown skull. Software developed by Venezis and colleagues enhanced the laser scanner system and allowed the addition of soft tissue features on a 2D file exported from their system (Vanezis and Vanezis 2000, Vanezis et al 2000). Figure 3.6 is an example of a facial approximation produced by this prototype.

The Computational Forensic Project was a pilot study conducted by the Oak Ridge National Labs (ORNL) and the University of Tennessee at Knoxville. The goal of this collaborative effort between forensic anthropology, computer science, radiology, and perceptual psychology was to develop a quick, cost-efficient, and reliable means of identifying victims. A database was formed from MRI data collected from European-American males between the ages 30 and 50. Statistical methods derived from artificial intelligence approaches were used for prediction of facial surfaces from bone surfaces. Marks et al 2000). An example of a facial approximation using this method is provided in Figure 3.7.

Volumetric data from CT scans of living subjects was used by Jones (2001) to compare features on a reference head to an unknown skull. The reference skull was warped over an unknown skull and an algorithm defined the tissue depths. Subjectivity was not eliminated as these initial systems used the existing tissue depth tables and operators subjectively chose the "reference" skull to be compared with the "unknown skull." (See Figure 3.8).

Vandermeulen and colleagues (2006) developed a concept for automated facial approximation by developing a database of three-dimensional cross-sectional CT images to determine distances between the skull and skin surfaces. Validation was

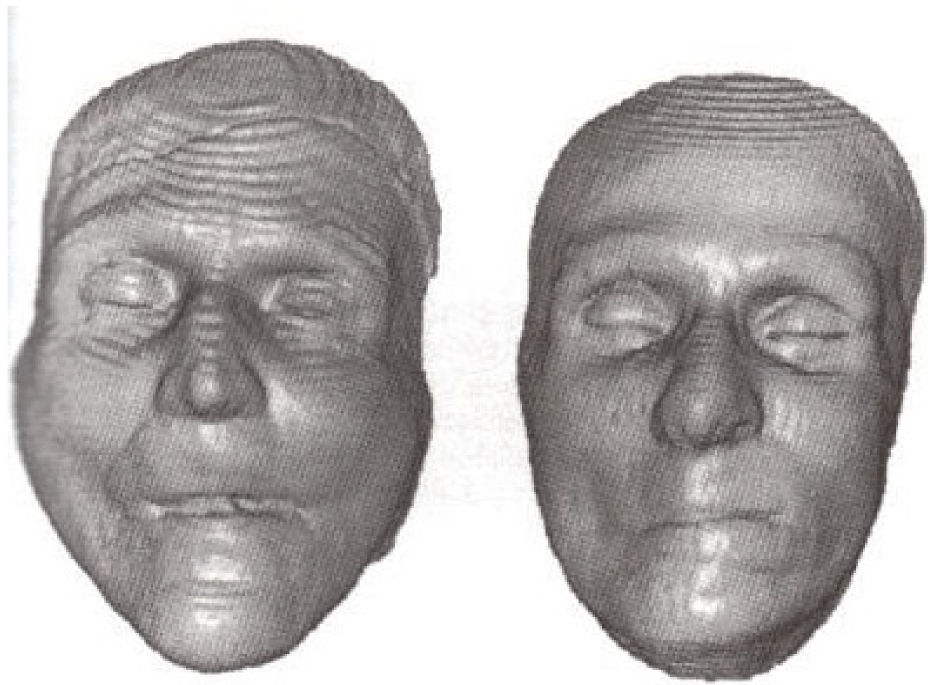


Figure 3.4: Computer facial approximation using volume deformation (left) compared to the actual death mask (right). (Quatrehomme et al 1997).

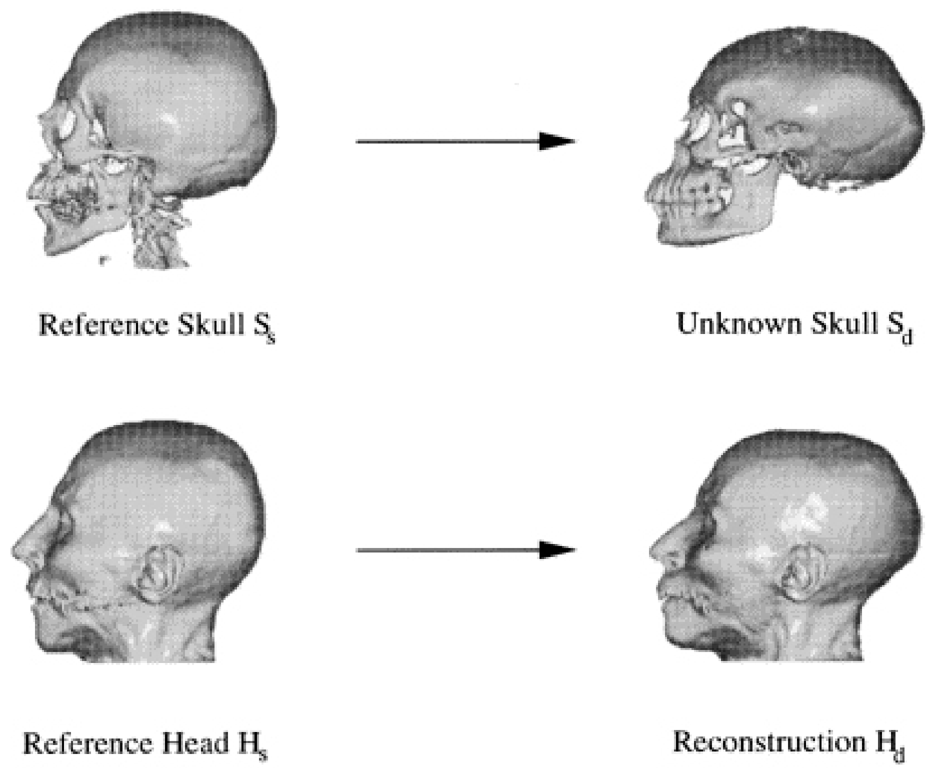


Figure 3.5: Computer facial approximation using volume deformation (Nelson and Michael 1998).

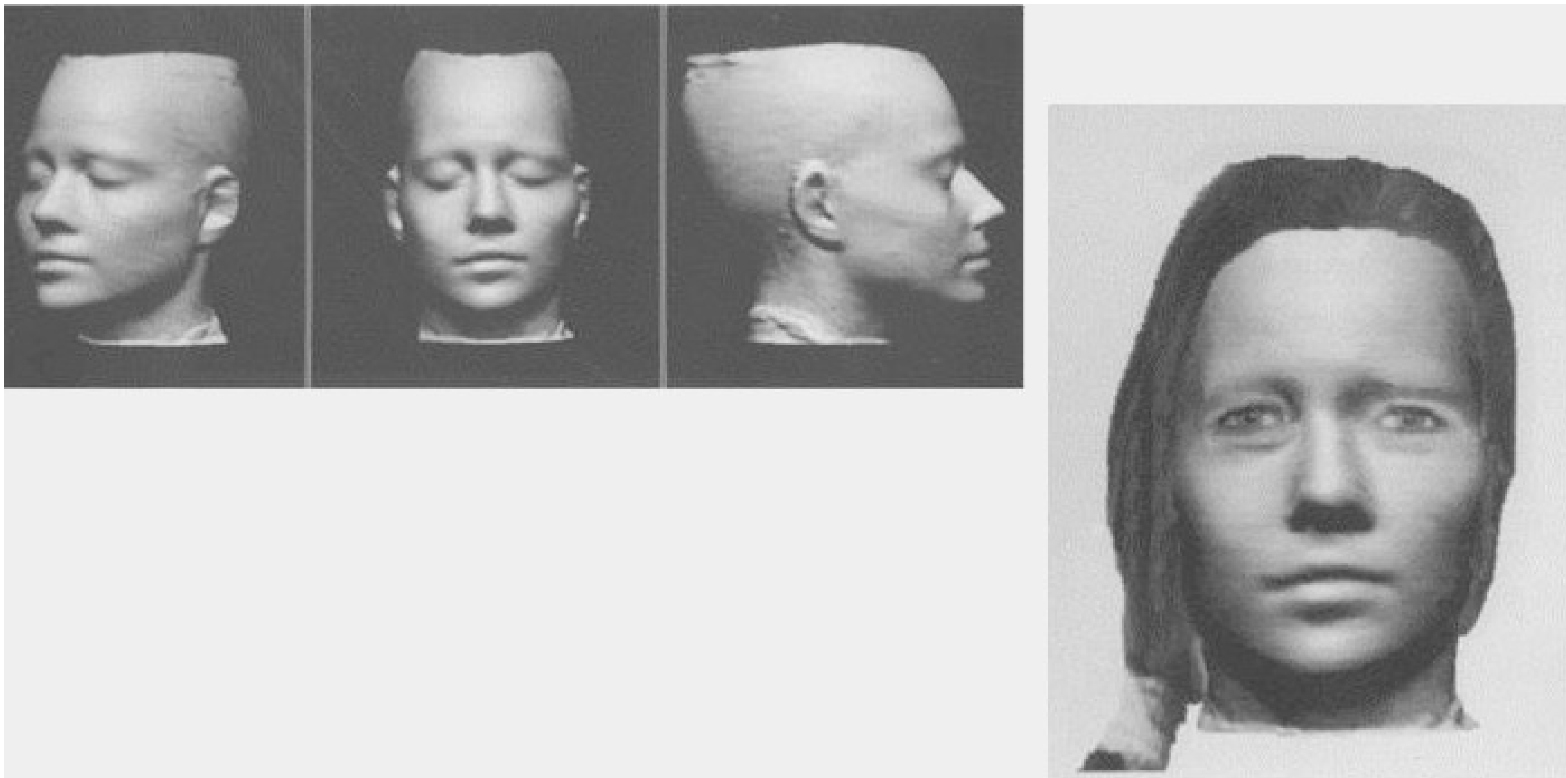
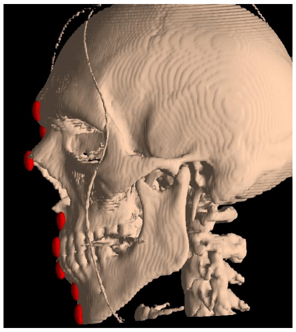
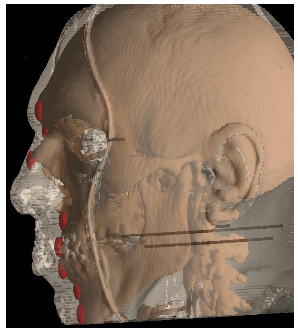


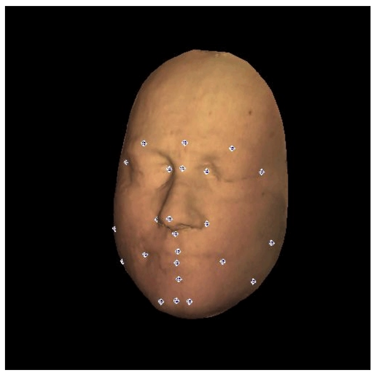
Figure 3.6: University of Glasgow Prototype for Computerized Facial Approximation (modified from Vanezis et al 2000).



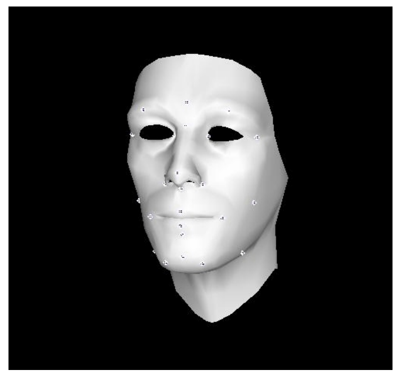
A. Skull with points selected.



B. Skull with transparent skin.

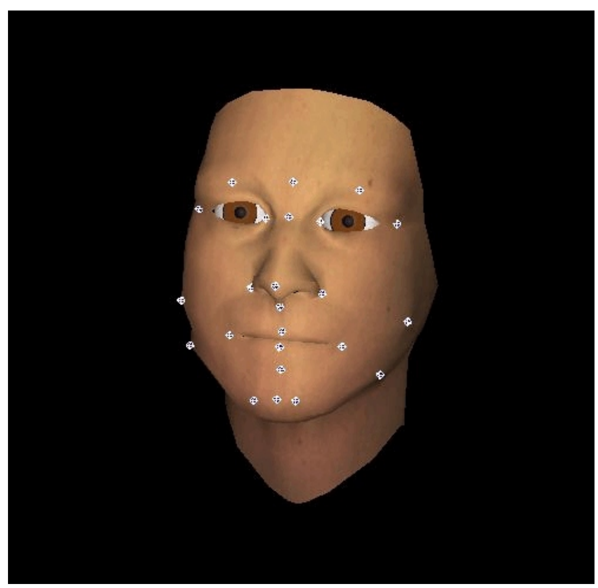


C. MRI with 24 key points.



D. Geometric Mask

Figure 3.7: University of Tennessee and ORNL (Project VIZ) (modified from Marks et al 2000).



E. Mask morphed to 24 points and eyes added.

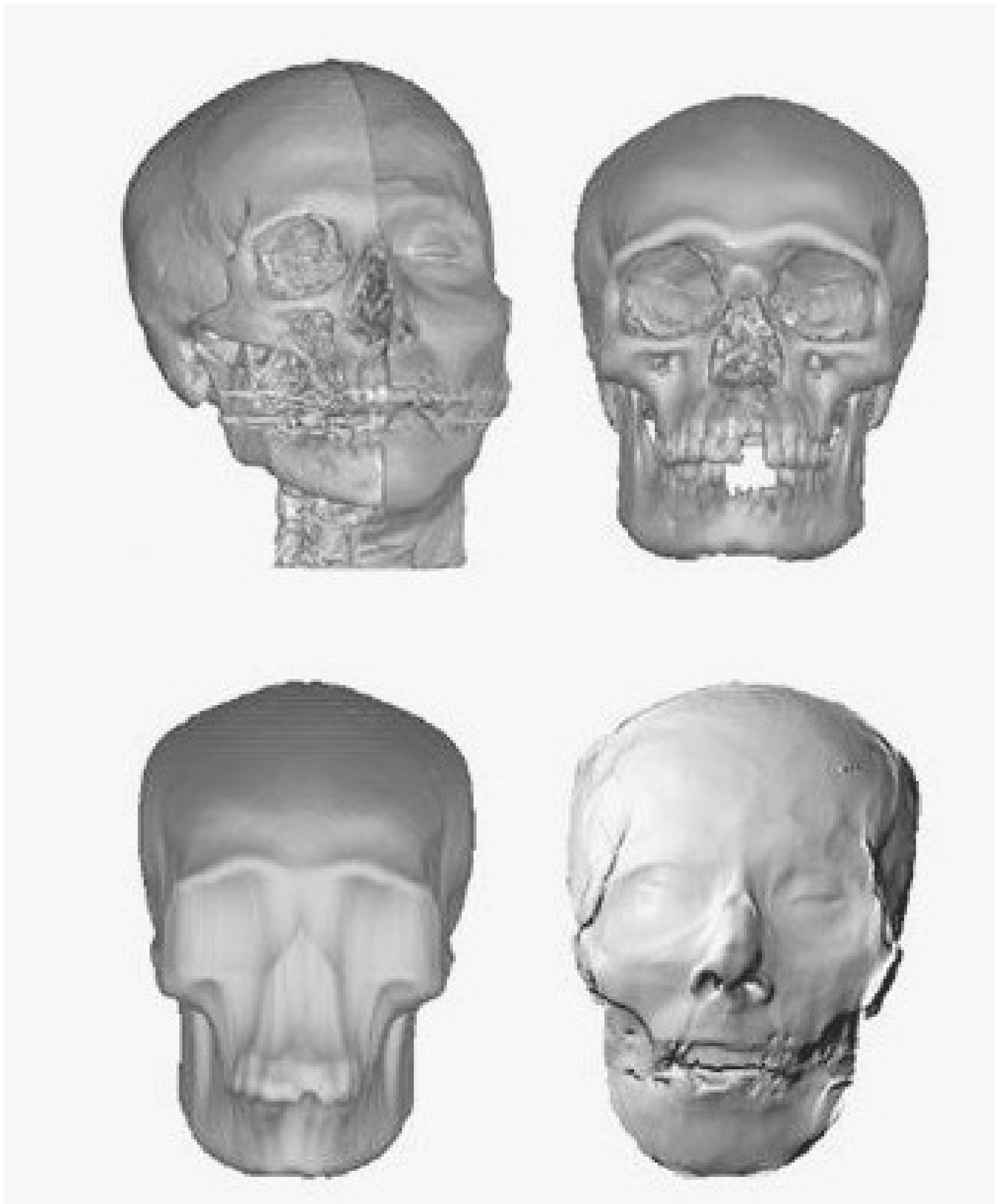


Figure 3.8: Facial approximation using volumetric data from CT scans (from Jones 2001).

done using a leave-one-out method, using a simulated (wet) skull from the database. Claes and colleagues (2006) tested this method on an actual human skull, using a database of CT scans and tissue depths from 118 living individuals. A statistical model of a face shape was combined with 52 soft tissue depths. Variations of facial approximations for a single skull were used to illustrate the ability of the method for adjustments of weight and aging. Two facial approximations using this prototype illustrated in Figure 3.9.

Each research study has incrementally moved forensic facial approximation toward an automated method. The use of actual human skulls for validation testing (as opposed to simulated skulls from a database) has been infrequent. (Only the model developed by Vanezis and colleagues (2000) appears to have been used in forensic casework.) However, none were found to have tested the prototype accuracy by comparison of facial approximations created from human skulls with antemortem photographs.

The subject of this research, ReFace, introduced a novel method of automated skull registration for forensic facial approximations in 2000 (Tu et al 2000) at the 9th Biennial Scientific Meeting of the International Association for Craniofacial Identification. This new approach developed a statistical model from a dense set of data extracted from a database of CT scans of living subjects using principle component analysis (PCA). A reconstruction algorithm was used to extract a 3D model of a unknown skull from a CT scan, and an average estimate of the face can be completed in a few hours. (For detailed descriptions of the internal functioning of this prototype, see Tu et al 2000, Tu et al 2003, Tu et al 2005, Miller et al 2003, and Turner et al 2005).

Facial Perception and Recognition

Extensive research has been done by cognitive psychologists on the ability of humans to perceive and recognize faces, and the importance of this research in the progress of forensic facial approximations is beginning to be recognized (Vanezis and Vanezis 2000, Clement and Marks 2005, DeGreef and Willems 2005, Wilkinson et al 2006).

Research designed to test the accuracy of manual or computerized facial approximations use images that are unfamiliar to research participants. Forensic applications of facial approximation depend on recognition of the “face” by someone familiar with the unidentified person. Psychological research indicates that people do very well in the identification of familiar faces, but very poorly in the identification of unfamiliar faces (Hancock et al 2000, Burton et al 2005). Bruce (1982) theorizes this is because the familiar face is already represented structurally and semantically in long term memory. Hancock (2000) contends that when confronted with an unfamiliar face, it is our perception of a face that determines how likely we are to remember it.

The history of facial approximation has attempted to predict the appearance of a specific individual by using “averages” of tissue depths that are known to vary between individuals as well as across populations. Psychological research indicates that distinct faces are more recognizable and better remembered than typical or “average” faces (Bartlett et al 1984, Hancock et al 2000).

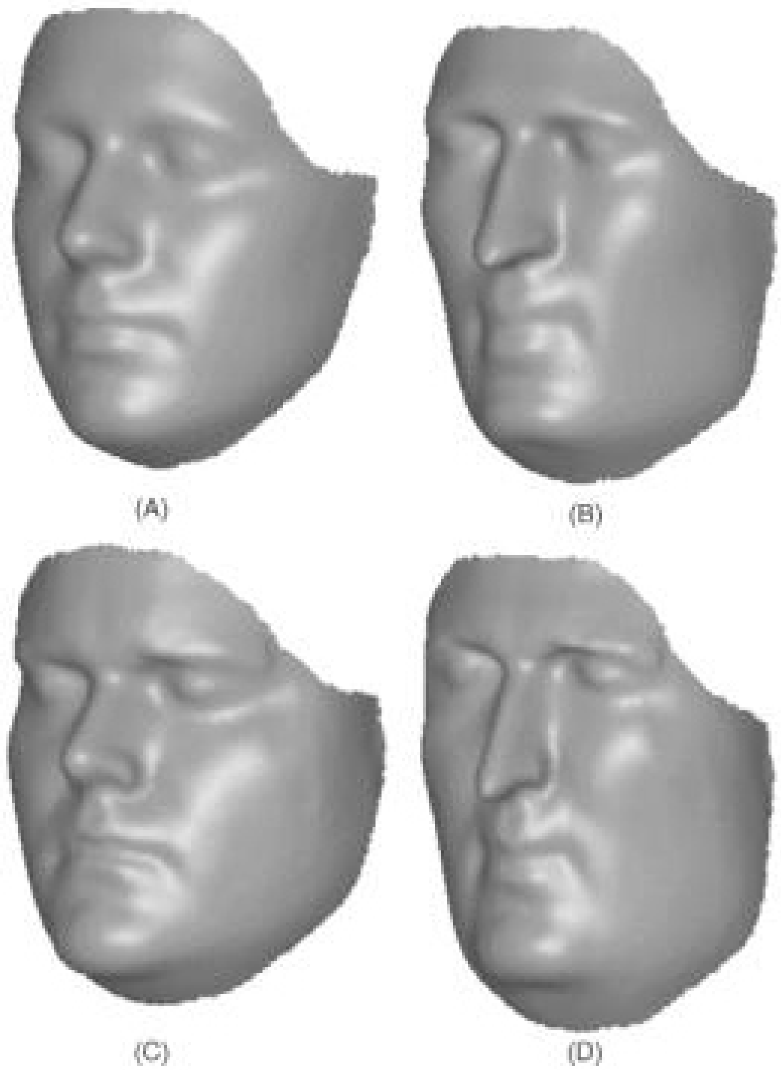


Figure 3.9: Computer facial approximations using a human skull and the resulting facial approximations. Face A is an average of faces in the database and Face B is the facial approximation. Face C is a specific facial template and Face D is the resulting facial approximation. (From Claes et al 2006).

The individual identity of a specific face has also been theorized to be a function of factors that can be measured and quantified, such as scale, position, and the ratio of facial features to one another (DeGreef and Willems 2005). This would seem to imply that all individuals presented with the same unknown face would process this information in the same way.

While recognizing that the anatomical configuration and specific facial features are important in the recognition of faces, cognitive psychologists have developed theories of facial recognition that emphasize individual variation in the ability to recognize familiar and unfamiliar faces (Doty 1998). Facial prototype theory contends that the ability of individuals to recognize faces is due to mentally stored facial prototypes. In the global prototype theory, individuals identify a face by comparing it to a single mentally stored facial prototype, which is an average of all the faces ever encountered by an individual. Each new face is incorporated into the facial prototype, and thus the prototype is altered with every new face encountered by an individual. (Valentine and Bruce 1986, Bruce 1988, Doty 1998). A variation is the local prototype theory, wherein multiple facial prototypes are developed to recognize faces of different groups (Doty 1998). Psychological research has found a lowered ability to recognize variation in other population groups (Bruce 1986, Doty 1998), yet it was also found that repeated exposure to other population groups improve recognition of that group (Doty 1998).

Future Development

Previous attempts to fully computerize facial approximations have often not used the experience, knowledge, and skills of forensic artists (dismissing them as unscientific). Forensic artists do not typically have physical anthropological training to interpret landmarks, features, and pathologies of bone. Artists and scientists have different areas of expertise, and both have valuable contributions to make in the area of facial approximation. ReFace was conceptually designed by forensic artists, created and brought into reality by computer programers, and assisted in the development phase by forensic anthropologists.

More accurate tissue depth measurements with larger samples of living persons are available. However, with the possibility of using tissue depths *over all the points on the skull*, the continued practice of limiting tissue depth measurements to specific craniometric landmarks may not be the best method of computerized facial approximation systems. Since technology has provided the ability to move beyond collecting measurements at traditional craniometric landmarks, using them may provide greater accuracy.

Research in the area of facial recognition indicates that cognitive psychology may also be useful in understanding how and why humans recognize each other. A successful computerized facial approximation method may require a multi-disciplinary, holistic approach to be successful. The continuing progress of this prototype is promising. The potential benefits an automated system of facial approximation could provide in the identification of unknown human remains is enormous; the continuing progress of ReFace towards this goal is promising.

Chapter 4
Overview of Prototype (ReFace) Validation

Development and Validation History

In July of 2000, the Federal Bureau of Investigation (FBI) hosted the 9th Biennial Scientific Meeting of the International Association for Craniofacial Identification held in Washington, DC. The initial concept for ReFace was presented at this conference (Tu et al 2000). At the 3rd European Academy of Forensic Science Meeting in Istanbul, Turkey in September 2003, the FBI and General Electric Company (GE) announced the joint development of a computer program for facial approximation. The program used flesh deformation modes and a database of CT scans of living subjects. The collection of a planned database of 300 CT scans of living persons was in progress (Tu et al 2003, Miller et al 2003). The prototype was given the name ReFace in 2004, and the novel method of automated skull registration was described and operationally defined (Turner et al 2005). This automated process did not require the manual placement of tissue depth markers, as it relied on tissue depth information mathematically extracted from CT data.

Validation testing of ReFace with human skulls began in May of 2005. Through the Federal Bureau of Investigation's Visiting Scientist Program, an educational opportunity administered by the Oak Ridge Institute of Science and Education (ORISE), the researcher conducted an independent validation of this software in two phases. In Phase 1, the performance of the software was evaluated; in Phase 2, the accuracy of the software to produce accurate facial approximations from human skulls.

ReFace Database and Research Test Set

The projected ReFace living subject database was 300 CT scans of living individuals to included a range of age, sex, and ancestry (Tu et al 2005). It was theorized that the larger the subject pool of scans, the more accurate the facial approximation. However, actual testing of the prototype indicated little change in the facial approximation after 30 faces have been mapped. The current database (Beta Version 7.0) contains more than 250 CT scans of living subjects, and is the largest CT dataset of living persons found in existing literature. CT scans obtain depth data for the entire face in vivo, resulting in the actual skin-to-bone ratios for each subject in the database. ReFace uses this *entire* dataset, rather than specifically defined craniometric points.

ReFace Database Statistics (Beta 5.2)
The ReFace database in Beta Version 5.2 consisted of 218 CT scans of male and female living subjects. Subjects were all Americans from three ancestral groups: European (49 males, 50 females); African (15 males, 10 females); and Asian (47 females, 47 males). The age distribution range was from 18 to 60 years, with a mean of 35.42 (see Table 4.1 and Figure 4.1).

Table 4.1. University of Tennessee Validation Subjects by Age, Sex, and Ancestry

Age Distribution	19 or Less	20-29	30-39	40-49	50-59	60-69	70+	N =
Black Males	0	1	1	0	0	0	1	3
Black Females	0	1	0	0	0	0	0	1
White Males	1	2	2	2	5	7	11	30
White Females	3	2	0	2	5	5	2	19
Total by Age	**4**	**6**	**3**	**4**	**10**	**12**	**14**	**53**

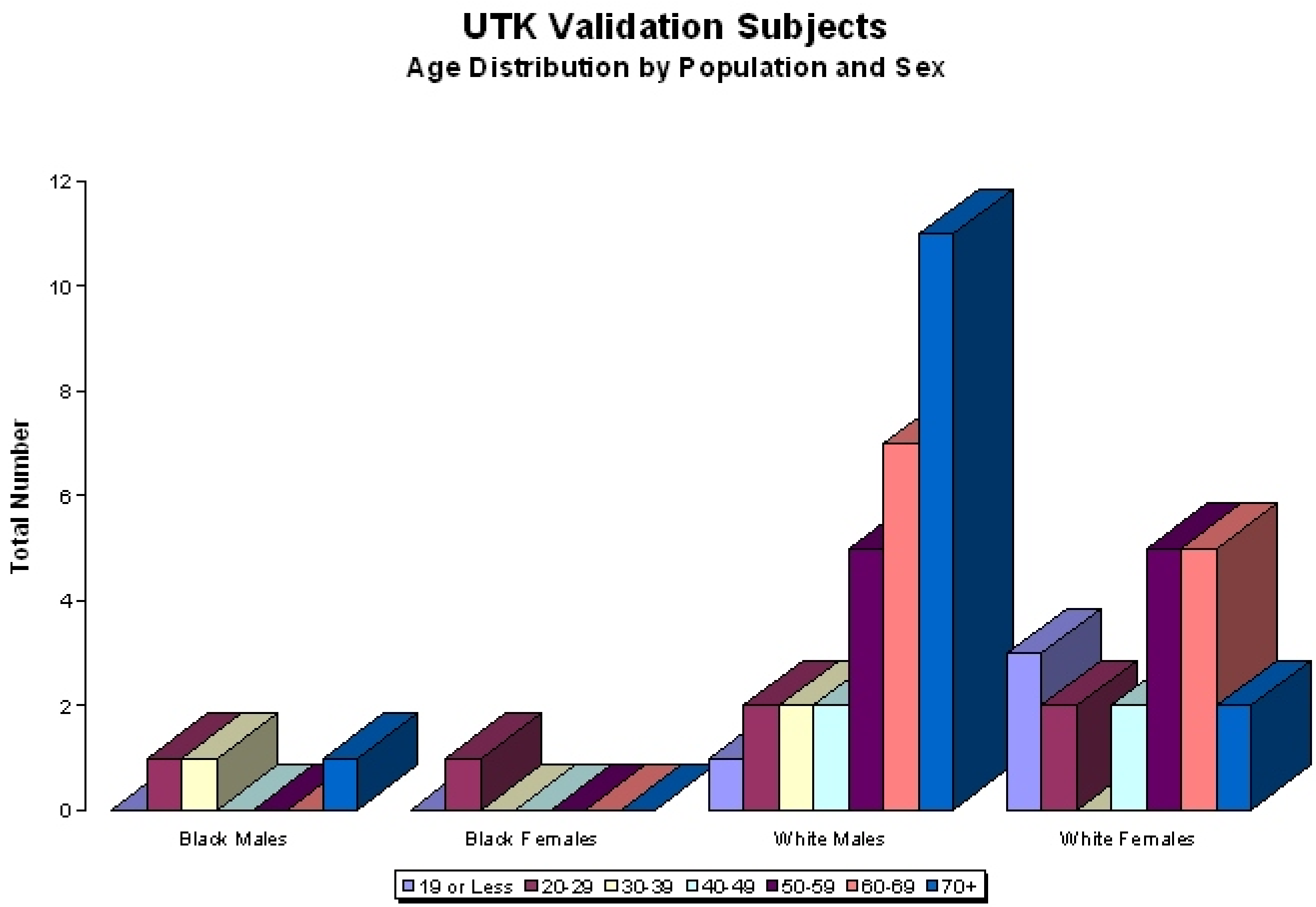


Figure 4.1 University of Tennessee *Phase 1* Testing Sample by Age, Sex, and Ancestry.

Research Test Set Statistics
The research test set consists of CT scans of 53 articulated human skulls from the William M. Bass Donated and Forensic Skeletal Collections at the University of Tennessee in Knoxville. CT scans of this population consisted of American male and female skulls from two ancestral groups: European (30 males, 19 females) and African (three males, one female). Age distribution range was from 15 to 96 years, with a mean of 55.47 (see Table 4.2 and Figure 4.2).

Software Operation

The software operates in three distinct stages. CT data from an unknown skull is loaded into the program as a case file. A three-dimensional skull model is made by the prototype and stored as a file (bone.vtk). The program then uses the three-dimensional bone model as the foundation for the approximated face.

Facial Approximations by ReFace
The anthropology biological profile (sex, ancestry, age range) is entered into the program and a facial approximation is produced in approximately three hours (the actual processing time depends on the computer hardware). The program compares the unknown skull to the closest matches of sex, ancestry, and age from the database. The approximated face is then “morphed” over the unknown skull and an “average” facial approximation of the unknown skull is produced and exported to a file which can be viewed in three dimensions. The program default produces an approximation of average weight, although weight can be manipulated on a thin-heavy continuum for a “lineup” of approximated faces. The approximation can be exported to an output file (skin.vtk) which can be viewed in three dimensions.

Post-Processing Tools
Post-processing tools are in development that will allow an operator to adjust the shape of soft tissue features and to give the approximation a more lifelike appearance by adding skin texture, hair, and opening the eyes (Tu et al 2005). These features were not yet available for use in this research. During the preliminary photographic evaluation in Phase 2, a forensic artist “opened” the eyes and added hair to six facial approximations. The artist had not seen the antemortem photographs of the six subjects. Weight adjustments were subjectively made by the researcher based on viewing the antemortem photographs. These enhancements were done to illustrate the capability of the program. For the final *Phase 2* validation (Face Pool and Resemblance Rating), no weight adjustments were made.

Table 4.2: Reface Database by Age, Sex, and Ancestry.

Age Distribution	19 or Less	20-29	30-39	40-49	50-59	60-69	N
Black Males	4	9	1	1	0	0	15
Black Females	1	4	1	3	1	0	10
White Males	11	15	7	13	3	0	49
White Females	2	17	9	15	7	0	50
Asian Males	4	21	12	8	1	1	47
Asian Females	8	22	12	4	1	0	47
Total by Age	**30**	**75**	**42**	**44**	**13**	**1**	**218**

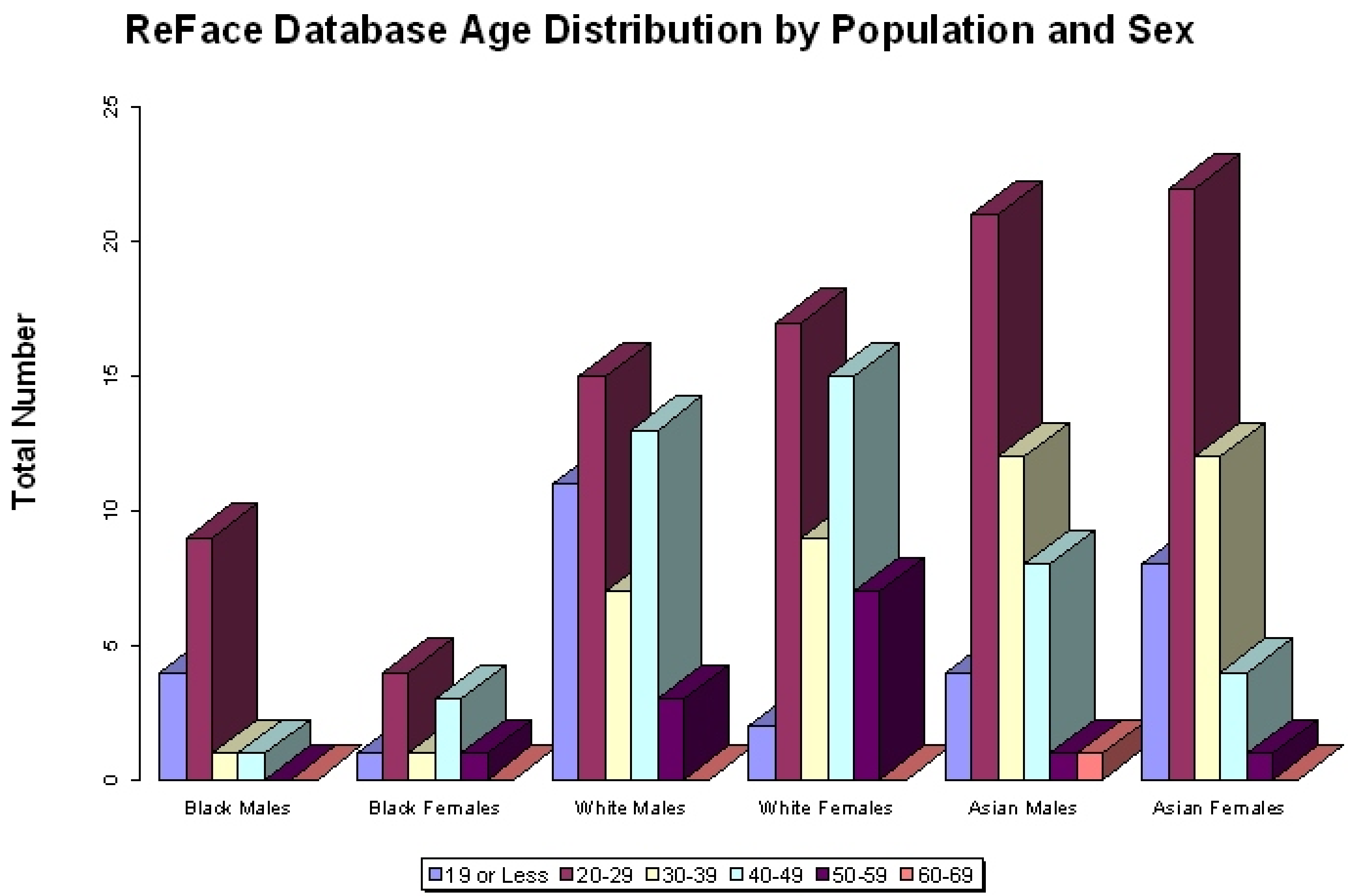


Figure 4.2: ReFace Database by Age, Sex, and Ancestry.

Chapter 5
Phase 1: ReFace Prototype Software Testing

Overview of Prototype Software Testing (*Phase 1*)

ReFace was evaluated in stages from the initial prototype through Beta Version 5.2. (Beta Version 6.0 was used in *Phase 2* validation testing, but was not evaluated.) Beta Versions 1.0 through 3.1 were tested with the initial sample of 30 University of Tennessee skulls and one wet skull (GE3) derived from the ReFace database of living subjects. Beta Version 4.0 was tested with four Bone Clones® replica skulls to evaluate articulation and packaging materials and to obtain CT scans of replica skulls of Asians for testing. Beta Version 5.0 and 5.1 were tested with the full sample of 53 skulls and the GE3 sample wet skull. Beta Version 5.2 corrected a programming error with the approximations of African ancestry, and was tested with four African ancestry UTK sample skulls (three males and one female).

Methods and Materials: *Phase 1*

Selection of Subjects for Study

From 1971 to 2006, the University of Tennessee Anthropological Research Facility (ARF) received approximately 1000 donated human remains of known sex, age, and ancestry. Case files from the William M. Bass Donated Skeletal Collection ("Donated Collection"), and the William M. Bass Forensic Skeletal Collection ("Forensic Collection") at the University of Tennessee, were reviewed to identify cases with antemortem photographs available for comparison with facial approximations made by ReFace. The complete sample set consisted of 53 skulls of known age, sex, and ancestry, that had antemortem photographs for comparison with computer facial approximations generated by ReFace.

Identification of First Sample for Testing
Case files from the Donated Collection (1981 to 2004) were personally reviewed by the researcher, and 48 cases with antemortem photographs were identified: 16 were still being processed and 30 were available for this study. There were 20 males (19 White and one Black) and ten females (all White).

Identification of Second Sample for Testing
ARF provided the following information to the researcher. A computer search of available records from 2003 to 2006 identified 12 additional subjects with antemortem photographs from the Donated Collection. After physical examination of the potential subjects, 12 new sample cases were added (three White females and nine White males). To obtain additional age and ancestral diversity for the second sample, ARF conducted a computer search of available records from 1971 to 2005 from the Forensic Collection. Potential subjects had to be positively identified and have an antemortem photograph to be included in the test sample. Fifteen potential subjects were identified, two of which were eliminated due to facial trauma and one subject eliminated as the

identification was found to be probable and not positive. The search identified eleven new sample cases (two Black males, one Black female, two White males, and six White females).

Articulation, Packaging, and CT Scanning of UTK Test Skulls

Each sample skull was articulated by the researcher and transported to the University of Tennessee Medical Center for CT scans. All scans were made on a GE Light Speed CT System (16 slice, helical scan type, with 0.8 second rotation time).

First Scan Set (30 donated skulls)
The first set of UTK sample skulls was articulated with one inch squares of small bubble wrap or foam sheets. Wooden dowels were used for proper spacing of the mandible; they were inserted but not glued. Skulls were held together with either (1) thin rubber bands, or (2) strips of bubble wrap. Five or six skulls were packed in a long cardboard box with inserts for stabilization and protection during transport.

All subjects were scanned at 5.0 mm slice thickness. Individual skulls were placed horizontally on a Styrofoam ™ ring base. This position was the same as that of the living database subjects, and therefore would have the same slice orientation as the living subjects. Living subjects in the ReFace Database were scanned at either 5.0 mm or 2.5 mm slice thickness.

Second Scan Set (Donated: 11 Rescans, 12 New ; 11 New Forensic)
The second set of UTK sample skulls was articulated with either (1) cotton balls and epoxy, (2) museum putty, (3) museum gel, or (4) museum wax. Wooden dowels were used for spacing with edentulous skulls, and were glued in place with epoxy.

All subjects were scanned at 2.5 mm slice thickness. Additional scans were made of two subjects at 1.25 mm and at .625 mm slice thickness for comparison. Prior testing with replica skulls on an industrial CT machine indicated an estimated 0.2 mm scan resulted in a large increase in file size which significantly slowed processing time without a discernable improvement in the resulting facial approximation. Additionally, the rubber band used to hold the skull together (used in the first sample at 5.0 mm slice thickness) was visible at the higher resolution. Since it was unknown if the rubber bands would be visible or interfere in the CT imports at 2.5 mm, 1.25 mm, or .625 mm, the second set of skulls were not articulated in the same manner as the first set.

Transport of Subject Data
CT scans were downloaded to portable media and brought with the researcher to the Counterterrorism and Forensic Science Research Unit, FBI Laboratory Division, in Quantico, Virginia. These data files were used in the validation of the ReFace Prototype (*Phase 1* and *Phase 2* testing). Details of the validation testing are set forth in Appendix C.

Results of *Phase 1* Testing

CT scans of the subject skulls were used to test the software performance from the initial prototype through Beta Version 5.2. Details of the testing are provided in Appendix C. Results of the first set of scans were analyzed, and new materials were tested prior to the second set of scans. A revised protocol was developed and used with the second set of scans. Results of the second set of scans were analyzed, and a final protocol was developed for articulation and packaging of skulls for CT scanning, as well as the optimum slice thickness. This material is presented in Appendix D.
The following areas were identified as limitations for further testing:

Aging. There was no ability for aging subjects or for the approximation of subjects under age 18. As a result, approximations of subjects over age 45 appeared youthful and subjects under age 16 appeared older. It was unknown if the inability to age (older or younger) was the result of the software programming or the lack of subjects of comparable age in the database.

Hair and Eyes. Facial approximations have closed eyes, and are without head or facial hair. Whether these will detract or interfere with the ability of humans to cognitively compare the approximations with antemortem photographs was unknown. The addition of hair and the opening of eyes were in development. In the interim, these features can be added by forensic artists to the completed computer approximation. Such artistic additions were made on six computer generated approximations for the *Phase 2* Preliminary Photograph Validation (described and illustrated in Chapter 6).

Weight. The prototype defaults to average weight, but adjustments for variations in weight (from heavy to thin) can be made by an operator by increasing or decreasing (from -9.0 to +9.0), which represent two standard deviations from the average reconstruction. The first two Effects appear to consistently increase or decrease weight. The adjustment is a manual, subjective process done by an operator after the reconstruction has been completed.

Dental Noise. The most severe program limitation is the inability of the prototype to build an accurate facial approximation on any skull with metal dental artifacts, such as metal dental fillings, clasps, screws, crowns, and metal based dentures (partial or complete). The metal creates a flash on the CT DICOM images, which the program incorporates as bone. The resulting bone model has spikes of bone extending horizontally from the lip and cheek areas. Facial approximations over these spiked bone models are deformed in the lip and cheek areas. This single program limitation eliminated many University of Tennessee test skulls from the validation study. As it can be expected that forensic skulls will have some type of dental work that includes metal, this issue must be resolved before ReFace can become fully functional.

Elimination of Subjects from Testing

Subjects were eliminated from *Phase 1* of the validation testing due to sample limitations, program limitations, program errors, or a combination of more than one category. Eight subjects were eliminated due to sample limitations: CT scan errors (5), articulation errors (1), poor quality photographs (2). Twenty three University of Tennessee sample subjects were over age 60, and eliminated due to the inability of the program to age. Elimination of subjects due to program limitations (Beta 5.2) are: 23

subjects over age 60; all Americans of African ancestry (insufficient database), and 11 subjects with metal dental artifacts.

Ten additional test subjects were eliminated due to program errors of unknown cause. New problem areas developed after installation of Beta Versions 5.1/5.2. New reconstructions had inexplicable extensions of nasal bones and exposed tips of mandibles (chin). Some reconstructions had both. Together with programming limitations (inability to age, bone model dental spikes), and subject limitations (CT errors, poor quality photographs, weight adjustments needed), there were only three subjects of the original pool of 53 that met ideal criteria for *Phase 2* photographic validation, with 17 potential subjects.

Chapter 6
Phase 2: ReFace Prototype Preliminary Validation

Introduction

A preliminary photographic validation was designed with ReFace Beta Version 5.2 and presented at the 12th Scientific Meeting of the International Association for Craniofacial Identification Abstracts (IACI) in Istanbul, Turkey (November 1-4, 2006).

Methods and Materials

ReFace Beta Version 5.2
Facial approximation were prepared with ReFace Beta Version 5.2. There were 99 subjects (49 males and 50 females) in the ReFace database of Americans of European descent. Age distribution for males was 18 to 54, with a mean of 30; age distribution for females was 18 to 58, with a mean of 35.

Criteria for Selection of Subjects for Photographic Comparison
For this presentation, six White subjects (Americans of European ancestry) were selected (three males and three females) for comparison with antemortem photographs. This selection provided a representative sampling of thin, average, and heavy weights (see Figure 6.1); dentition (edentulous, partial dentition, and complete dentition); and a sample of a proposed protocol for the correction of CT DICOM images to compensate for image distortion due to metal dental artifacts (see Figure 6.2).

Previous problems identified with the prototype included the inability of the software to age, to add hair, or to open the eyes. For this presentation, only subjects that were age 60 or less, (or whose antemortem photograph was of a younger age) were selected for facial approximations. For the comparison with antemortem photographs, weight selection was made for subjects who appeared very thin or very heavy. Decisions to increase or decrease weight were subjectively determined by the researcher after viewing the antemortem photographs.

Target Subjects

<u>Male Subject 1: (University of Tennessee 02-89D)</u>
White male, age 36, 220 pounds, height of 72 inches. The antemortem photograph was an actual color mug shot that was scanned and the face cropped. When the photograph was enlarged to include in the face pool, the resulting resolution was less than optimal. The target subject's age in the photograph was 32 and he appeared to be of lower weight than the weight at death. The CT slice thickness was 2.5 mm. This subject was selected to represent average weight and partial dentition.

<u>Male Subject 2: (UTK 24-04D)</u>
White male, age 61, 210 pounds, height unknown. The antemortem photograph was a color candid that was out of focus and of low resolution. The CT slice thickness was 5.0 mm. This subject was selected to represent edentulous and heavy weight.

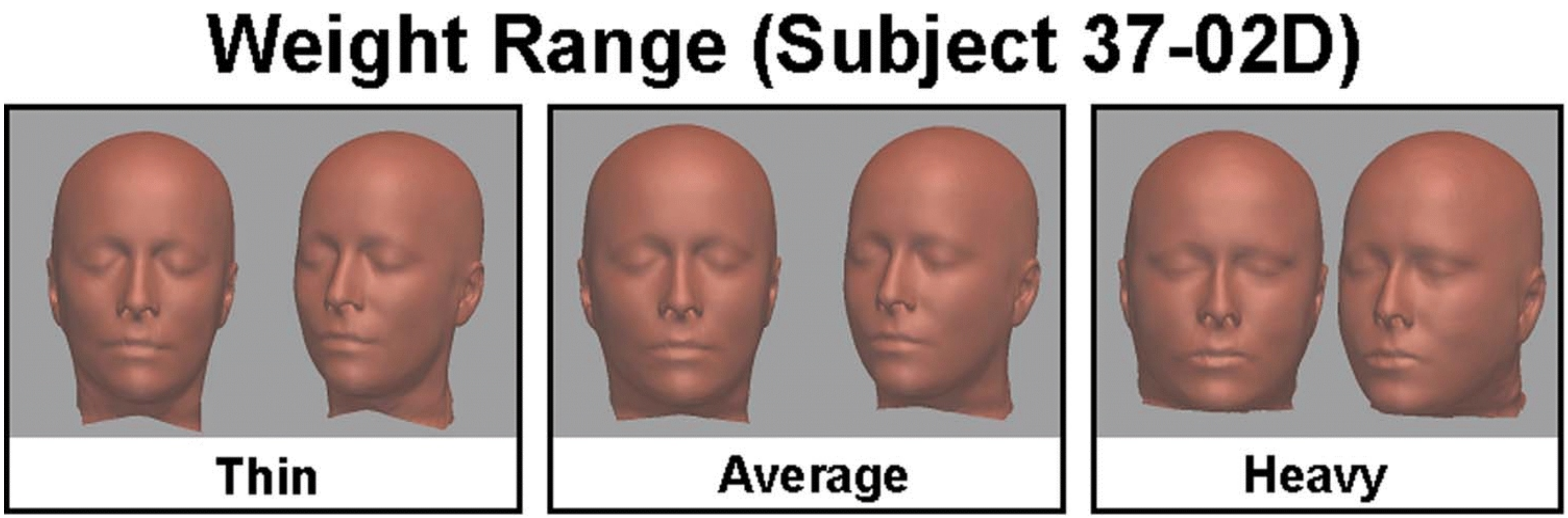


Figure 6.1: Facial approximation weight range of thin, average, and heavy for Subject 37-02D.

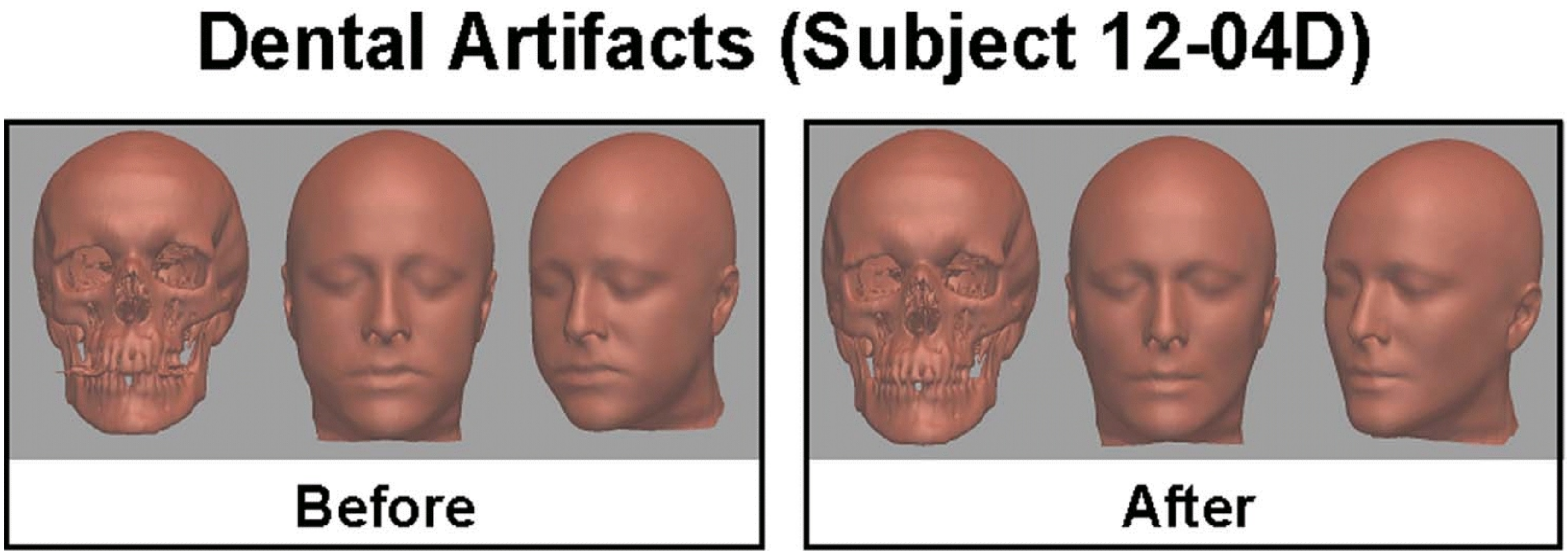


Figure 6.2: Dental artifacts image distortion (before) and DICOM corrected (after).

Male Subject 3 (UTK 06-04D)
White male, age 16, 50 pounds, height unknown. The antemortem photograph is a color candid that was slightly out of focus and of average resolution. The CT slice thickness was 5.0 mm. This subject was selected to represent thin weight and full dentition.

Female Subject 1 (UTK 12-04D)
White female, age 60, 125 pounds, height of 66 inches. The antemortem photograph was a black and white candid that was blurred and of low resolution. The CT slice thickness was 2.5 mm. This subject was selected to represent thin weight, full dentition, and to demonstrate a DICOM adjustment protocol developed by the researcher.

Female Subject 2 (UTK 37-02D)
White female, age 52, 183 pounds, height of 64 inches. The antemortem photograph was a color candid of average resolution. The CT slice thickness was 5.0 mm. This subject was selected to represent heavy weight, partial dentition, and to demonstrate a DICOM adjustment protocol developed by the researcher.

Female Subject 3 (UTK 03-01D)
White female, age 62, 104 pounds, height unknown. The antemortem photograph was a black and white candid of good resolution, with the subject much younger and of increased weight. The CT slice thickness was 2.5 mm. This subject was selected to represent average weight and edentulous.

Presentation of Subjects
Completely automated facial approximations (without hair and with closed eyes) were presented for comparison with antemortem photographs. To assist in the visual comparison, a forensic artist added hair and open eyes to computer approximations of these six subjects. For comparison with antemortem photographs, weight selection was made for subjects who appeared very thin or very heavy in their antemortem photographs. Facial approximations were aligned in approximately the same orientation as the antemortem photographs.

Results of Preliminary *Phase 2* Testing

Evolution of the prototype from Beta Version 1.0 to Beta 5.2 was illustrated for the same subject (06-04D) in Figure 6.3. This facial approximation shows the program default of average weight. The same subject is included in a photographic comparison showing an extreme reduction for weight (subject was approximately 60 pounds at death due to disease, and was also very thin in the antemortem photograph).

The photographic lineup in Figure 6.4 was presented for review and comment from conference attendees. Comments were frequent and positive as to the resemblance between the computer images and the antemortem photographs. Under optimal conditions, the results of facial approximations completed by Beta Version 5.2 bear a striking resemblance to antemortem photographs.

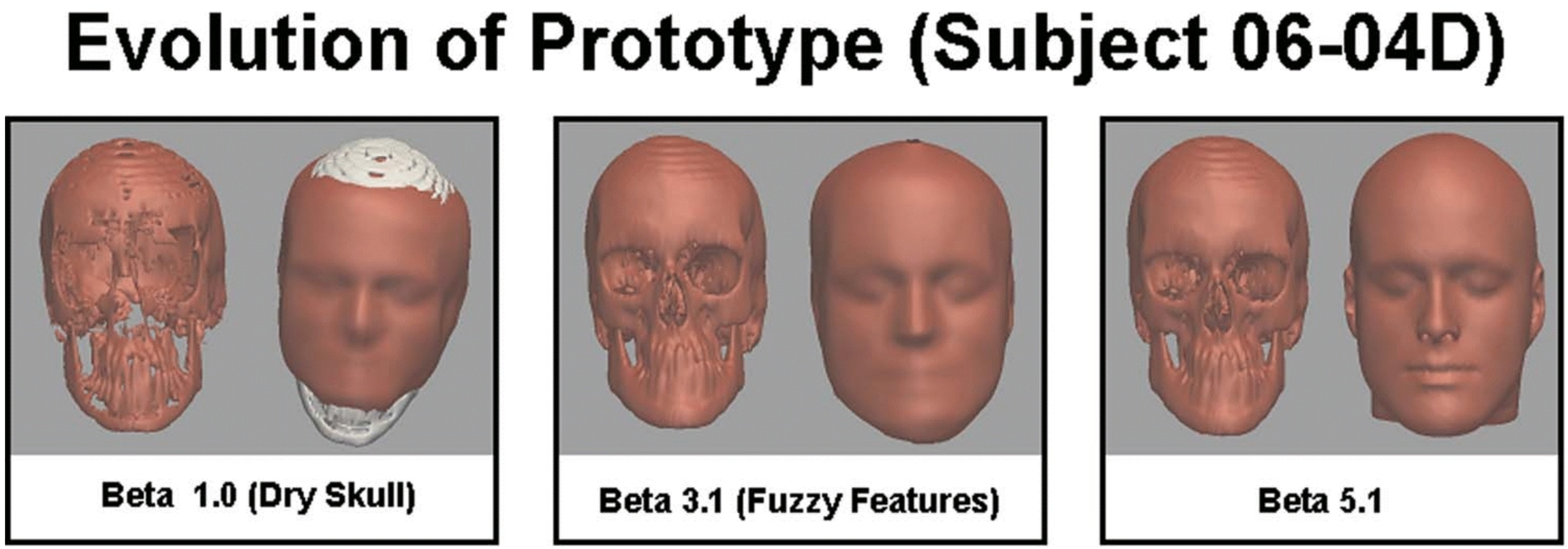


Figure 6.3: Bone and Skin Models from ReFace Beta Approximations (Versions 1.0, 3.1, and 5.1).

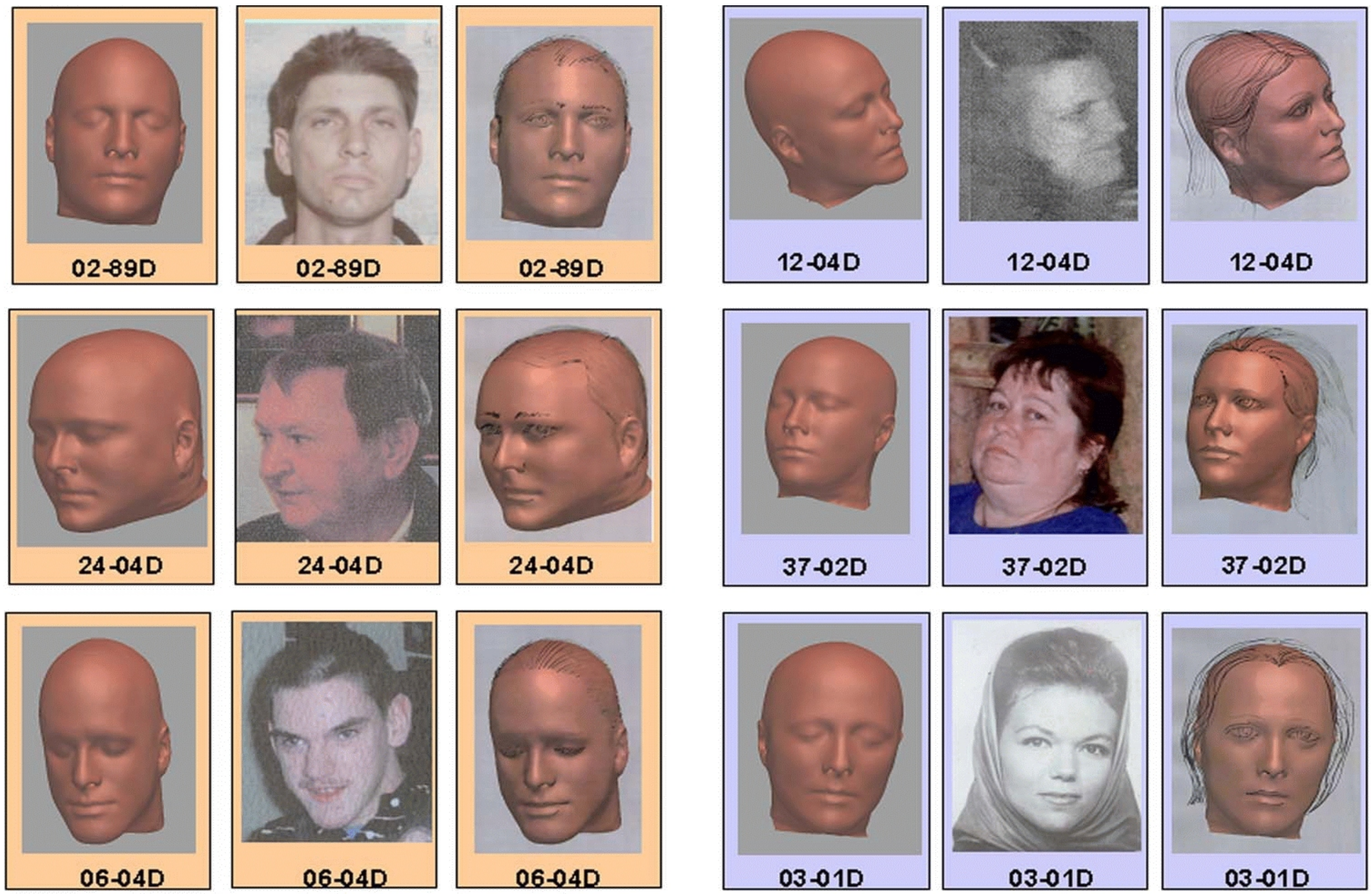


Figure 6.4: Comparison of ReFace approximations with antemortem photographs and with artistic addition of hair and open eyes, (Artist Myke Taister, reprinted with permission).

The issues of time, expense, and subjectivity have been addressed by this prototype: (1) *Subjectivity*. Beta Version 5.2 was fully automated with no subjective involvement by the operator, (2) *Time*. Facial approximations are completed in less than three hours, and (3) *Expense*. Although the expertise of forensic anthropologists and forensic artists were included in the development of this prototype, the operation of the software is not anticipated to require either.

Chapter 7
Development of Research Design for Final *Phase 2* Validation

Introduction

A review of existing literature (English only) did not find any published studies that specifically compared a *computerized* facial approximation with antemortem photographs or a photographic "face pool." Pioneer research comparing facial approximation with death masks or postmortem photographs, such as Von Eggeling (1913), Stadtmüller (1922), Diedrich (1926), and Gerassimow (1955, 1968) were not used in this research design (see Table 1, page 230 in Helmer et al 1993).

Although a single research project that compared computer images to computer images was found (Wilkinson et al 2006), this research validation project appears to be the first known comparison test of computer generated facial approximations with antemortem photographs. Key research articles used for this research design are summarized below.

Summary of Key Research Articles

Snow, Gatliff, and McWilliams (1970) compared clay reconstructions with antemortem photographs in face pools of seven photographs. Results for the female subject (N = 104) were 11% over random chance (26% correct), and results for the male subject (N = 200) were 54% over random chance (67% correct). Police officers and civilian females scored higher than civilian males.

Haglund and Reay (1991) enlisted nine different forensic artists who prepared 24 facial approximations in an attempt to identify victims of the Green River Serial Murder case. Depending on their preference, artists were provided with the actual skulls, models of the skull, or photographs or radiographs of the skull. Reconstructions by different artists on the same skull varied considerably, but some proved to be quite accurate when compared to antemortem photographs of subsequently identified victims.

Helmer and colleagues (1993) used two different examiners for the construction of 24 facial reconstructions (12 each), which were then rated by three different examiners on a resemblance rating scale of 1 to 5. A comparison of the reconstructions with each other showed 50% as approximate resemblance and 33% as close resemblance. Comparison of the reconstructions with antemortem photographs with the reconstructions (N = 24) were 38% close resemblance, 17% approximate resemblance, 42% slight resemblance.

Stephan and Henneberg (2001) designed an elaborate test using four skulls, four different approximation techniques, sixteen approximations, 37 assessors, and a face pool of ten photographs. Only one facial approximation resulted in true positive identification rates above random chance at statistically significant levels. Such low ratings appear to be unique in the literature, and may be more reflective of this research design, rather than the accuracy of facial approximation.

Wilkinson and Whittaker (2002) produced five traditional clay reconstructions on five female juvenile skulls. The five target subjects were included in a face pool of 10 photographs and all five were the most frequently identified in the face pool. Results for the face pool test (N = 50) were a mean hit rate of 34% above random chance (44% correct). This study also used a five-point resemblance rating scale and all reconstructions were rated as close resemblance to the target individuals.

Stephan and Arthur (2006) prepared two approximations from the same skull to test the accuracy of resemblance rating and recognition rating (face pool) tests. One reconstruction was done by an experienced practitioner who had access to the antemortem photograph of the subject; the second reconstruction was done by a "novice" without access to the antemortem photograph. The authors contend the resemblance ratings for both are similar, even though their own analysis indicates a statistically significant difference ($p < 0.05$) between the two scores. Face pool results show a 98% correct response for the experienced practitioner with access to the antemortem photograph and a 12% correct response for the novice without access.

The unique project of Wilkinson and colleagues (2006) produced two computer assisted facial approximations that were derived from CT scans of living individuals from the ReFace database. The approximations were then compared (N = 52) in face pools of five *computer* surface model images. The combined hit rate was 50% above random chance (70% correct).

Based upon this review, the methods selected for the final photographic validation were (1) Face Pools and (2) Resemblance Ratings.

<u>Resemblance Rating Scale</u>
Resemblance rating scales were generally on a Likert scale of 1 to 5 with categories of none or no resemblance to strong or high resemblance. The resemblance rating chosen for this research was derived from Helmer and colleagues (1993) with the single variation that the numerical descriptions were reversed, with 1 as no resemblance and 5 as high resemblance. Resemblance rating scales were generally from 1 to 5 with descriptive categories variations of the scale used by Helmer and colleagues (1993).

<u>Face Pool Criteria</u>
Different researchers used different numbers of photographs in their face pools, but did not indicate the criteria upon which the number was selected. The number of photographs ranged from a minimum of 5 to a maximum of 10. The face pool design used in this research was based on previous research and recommendations in law enforcement training materials for eyewitness identifications.

Eyewitness Identifications by Photographic Lineups
In law enforcement photo arrays, witnesses are presented with a group of photographs, which are shown either sequentially (one at a time) or simultaneously (as in a face pool). Eyewitnesses have a mental image of a person and are looking for a photograph that matches that mental image. Such identifications can be influenced by emotional and

psychological factors, and the consequences for a wrong identification may be severe. A research study by Doty (1998) found cross-cultural identifications were much less accurate in the identification of familiar subjects, which could lead to false identification. As identified suspects may be prosecuted for crimes, courts assess whether the design of the photo array was fair, and whether the witness was unduly influenced by suggestive comments.

In forensic facial approximations, people are similarly looking for a person they know or whose photograph they have already seen. However, in research testing, participants are asked to identify someone they do not know and have never before seen. The answers of the participants are anonymous, and there are no negative consequences to anyone for a wrong choice.

The following law enforcement guidelines for eliminating bias in photo arrays were used in the design and presentation of the face pools used in this research. ("Eyewitness.." 2006, "Eyewitness..." 1999, Friend 1991).

1. There should be a minimum six photographs, including the target individual.

2. None of the foil photographs should unduly stand out.

3. Photographs in the face pool should be of people who are reasonably uniform in age, height, weight, general appearance, and of the same sex and ancestry as the target individual.

4. Photographs should be of similar size and composition. Black and white photographs should not be mixed with color photographs.

5. The face pool should not include more than one photo of the target individual.

6. The face pool and comparison approximation should be shown to one person at a time; then removed from view before the next approximation and face pool are presented.

7. No suggestive statements should be made.

Selection of Photographs for Comparison Face Pool

<u>UTK Subject Antemortem Photographs</u>
Photographs of the target subjects varied in size, resolution, and quality. Most color photographs are candid snapshots, but a few are professional studio type photographs. Black and white photo images include newspaper clippings, photocopies of photographs, and flyers for missing persons. A few subjects have a range of ages represented by photographs, but most have a single photograph. The exact age of the subject at the time the photograph was taken is unknown. Most are posed in a three-quarter view and are smiling.

Photographic Foil Selection
Over eight hundred photographs were examined to obtain face pool photographs for comparison with the target subject photograph. These photographs were primarily mug shots and employee identification badge photographs. Most were full frontal poses, with profile views of the mug shots, and few had any facial expression. The quality of these photographs was significantly better than the UTK antemortem photographs.

Presentation of Photographs
Previous research indicated that photographs for the face pool photo array are commonly mounted or printed on posters. Participants are shown approximation photographs and asked to choose a face (photograph) from the face pool that most closely resembles the approximation photograph. The quality of the antemortem photographs was determined to be insufficient for enlargement for a poster or PowerPoint® presentation. Microsoft® PowerPoint® 2000 was used to create three different face pools for ten target subjects. Each participant received individual items for comparison.

Proposed Participant Groups
Previous research using face pools for comparison of traditional, non-computer generated facial approximations indicate that (1) experts have a higher hit rate in the selection of target individuals in face pools than non-experts, and (2) females have a higher hit rate than males. Assessors for this study would ideally include both males and females of different ages and levels of experience or familiarity with facial approximation.

Face Pool Design

Test 1: Face Pool
Purpose of Test
To assess the ability of humans to identify a target subject in a face pool of photographs from a computer generated facial approximation.

Face Pool Layout
Part A. Seven photographs (including target) were used for the photographic face pool. Antemortem photographs from University of Tennessee subjects that were not used in the validation were selected for the face pools as they were the most similar in size, resolution, and facial expression to the photographs of the target subject. Photographs were standardized in size, color (gray-scale), cropped around the face, and standardized in resolution and pose as much as possible. The actual age of the target subjects in the antemortem photographs was unknown.

Part B. ReFace computer facial approximation in three poses (front, 3/4 profile, and profile) were printed below the face pool. No cropping was done to the approximation photographs. The three-quarter view for each approximation was added as this view had been found advantageous in the matching of unfamiliar faces (Bruce et al 1987). Figure 7.1 illustrates the placement of the face pool (Part A) with the ReFace approximation (Part B).

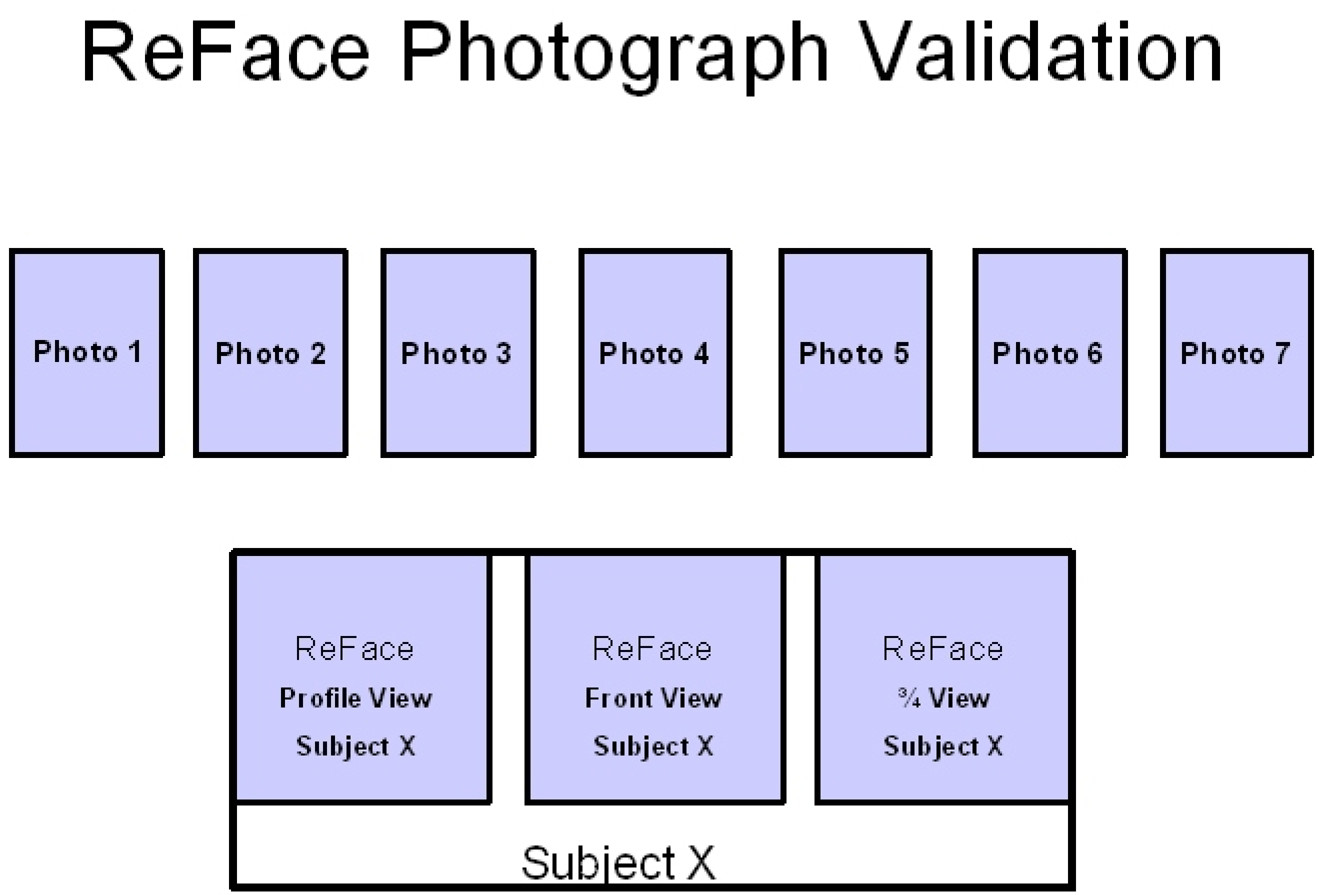


Figure 7.1: Design Layout for Photographic Face Pool Validation

Method of Selection
Assessors were requested to choose the photograph that *most resembled* the computer facial approximation. Implied, but not stated, was the possibility the target subject was not present in the face pool. However, no foil approximations were used in this test; all face pools contained the target subject.

Resemblance Rating Design
Test 2: Resemblance Rating

Purpose of Test
To compare a ReFace approximation of a target subject with the antemortem photograph of the target subject to assess the similarity between them.

Resemblance Test
Ten ReFace approximations were each aligned in approximately the same orientation as the antemortem photograph of the target subject. The assessors were asked to rate the resemblance of the facial approximation to the photograph on a scale of 1 to 5 (see Table 7.1). Figure 7.2 illustrates the resemblance rating design layout.

<u>University of Tennessee Subject Photographs</u>
Nine of the same target subjects from the Face Pool Test were used for the Resemblance Rating Test. Subject B was replaced in the Resemblance Rating Test with a foil approximation and a foil antemortem photograph (approximation and photograph were of different subjects).

Design of Validation Testing Kits

Groups of participants were available for testing in one hour time slots. Individual presentation of the approximations could not be accomplished in that time frame. A testing kit was designed to meet the group and time requirements while maintaining scientific integrity. The design was the same for both Face Pool and Resemblance Rating testing kits.

Randomization of Testing Materials

<u>Photographs</u>
Photographs were placed in individual envelopes and sealed. To randomize the presentation order, envelopes were divided into three piles, and then reassembled in random order. The envelopes were fanned out in alternating right-to-left and left-to-right order. The randomly assorted envelopes were placed inside a larger envelope with an answer sheet.

<u>Answer Sheets</u>
Answer sheets were sequentially numbered, divided into three piles, and then reassembled by sequentially taking one answer sheet from each pile.

Table 7.1: Validation Resemblance Rating Scale

Rating Scale		Degree of Resemblance
1	=	no resemblance
2	=	slight resemblance
3	=	approximate resemblance
4	=	close resemblance
5	=	strong resemblance

* [modified from Helmer, *et al* 1993]

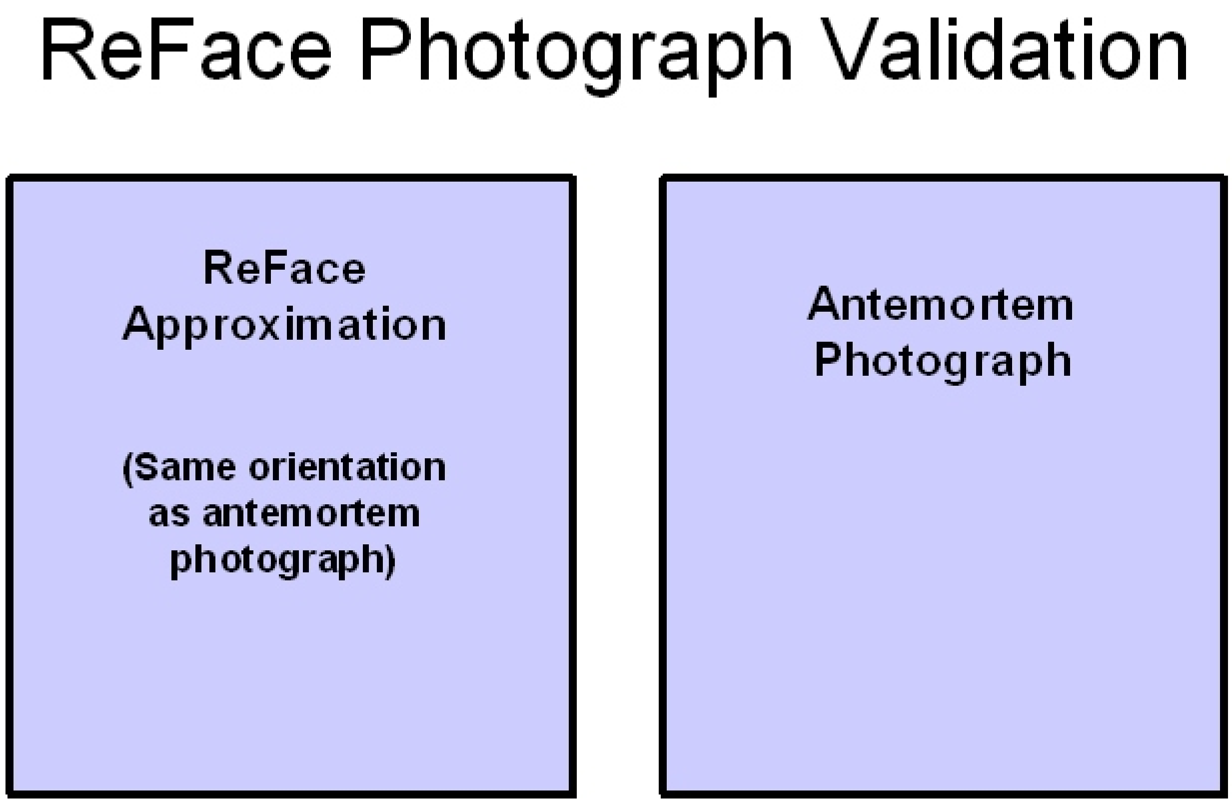


Figure 7.2: Design Layout for Photographic Resemblance Rating Validation.

Outside Envelopes
The outside envelopes were sealed. An instruction sheet was taped to the back of each envelope. A color-coded sheet with the test name was taped to the front of the envelope. Outside envelopes were randomly packed into a container for distribution to participants.

Distribution of Test to Participants
Validation ONE: Face Pool sealed envelopes were handed out to participants by two different administrators. No time limit was given to complete the first test. Participants were given the Validation TWO: Resemblance Rating envelope after completion of the first test.

Chapter 8
Phase 2: Final ReFace Prototype Validation

Introduction

The final *Phase 2* validation for comparison of facial approximations generated by ReFace to the antemortem photographs of test subjects was designed for Beta Version 5.2. However, of the initial sample of 53, only three subjects met all criteria for validation. Prior to the implementation of the *Phase 2* validation, Beta Version 6.0 was made available to the researcher. This version contained programming corrections that enabled the addition of seven additional target subjects (six met all criteria for validation; one subject had slight dental spikes in the skull model and was used to test the user tool for removal of extraneous material). Using the most current version of the software was deemed more useful to the FBI, and the researcher was instructed to use Beta Version 6.0 for the final validation. DICOM data from ten target subjects was loaded into ReFace and new facial approximations of ten target subjects were prepared using Beta Version 6.0.

Methods and Materials

Selection of Target Subjects

Selection of Target Subjects for Validation
Of the initial sample of 53, only three subjects met all criteria for photographic validation. However, there were 17 additional subjects with slight deviations from optimal criteria that might have been usable. The minimum sample size for validation with Beta Version 5.2 was three subjects and the maximum possible was 20 subjects. Beta Version 6.0 corrected the program errors of exposed chin, teeth, and nasal bones. Use of this version added seven additional subjects with optimal criteria, for a final test set of ten (10) subjects for the final photographic validation.

Target Subjects for Final Phase 2 Validation
Age, weight, and height were known for most of the target subjects. Initially, standardization of height-weight variables were to be used to determine whether the approximation should be adjusted for weight. However, visual examination of the antemortem photographs clearly indicated that some of the subjects were younger, and there was limited correlation of the weight at death with the weight in the photograph.

Selection of Face Pool (Non-Target) Photographs
Photographs were selected to create a group of cohorts of the same sex and approximate age. Five volunteers were individually shown prospective face pools (without an approximation), and were asked to choose a photograph that did not “fit” with the group. Selected photographs were removed from the pool and others substituted until the volunteers indicated that no one photograph stood out or seemed different from the others.

ReFace Prototype Beta Version (6.0)

Changes in Software Operation

Opening of the eyes had been partially implemented in Beta 6.0, but the development phase was not complete. A new user interface tool was introduced for the removal of dental spikes and extraneous material from the bone model, but this tool had not yet been tested. Adjustment of weight and adjustment of age were automated in this version, but had not yet been tested. No provision for the addition of hair was contained in this version. As a result, the approximations used in the *Phase 2* final validation still had closed eyes, were without hair, and there was no ability for aging.

Area of Potential Bias

Prior versions of the prototype were completely automated, and all imports of skulls and facial approximations were completed by the researcher. However, the user interface of Beta Version 6.0 required the manual placement of 19 points on specified craniometric landmarks around the eyes, nose, and mouth. Placement of points was done by moving colored computer “dots” in three-dimensional space using X, Y, and Z coordinates. In areas where the bone was missing, such as with edentulous skulls, the placement of points had to be subjectively estimated by the operator.

Change in Protocol to Address Potential Bias

There was no automatic override; the skull model required the manual placement of the landmarks for the completion of the skull model. To prevent researcher bias, an individual experienced in computer graphics placed the landmark “dots” using written instructions provided by the program developers. Once the skull models were completed, the researcher entered the age range and ancestry for each target subject, and the subsequent facial approximations were completed by the program.

Face Pool Photograph Selection

Antemortem photographs from the original sample of 53 University of Tennessee subjects were used for the face pools. Photographs were closely cropped around the face to standardize the size of the face. Color and black photographs were reduced to grayscale images.

Criteria for Target Subjects

Optimal criteria for selection of subjects for facial approximations were under age 60 (ideally age 25 to 35), with no metal dental work, antemortem photograph of good resolution, and of average weight, and a CT slice thickness of 2.5 mm. Table 9.1 provides a summary of the University of Tennessee target subjects selected for the final validation testing.

Subject A (UTK 11-04D)

White female, age 54, 120 pounds, height of 65 inches, edentulous. Photograph 6 in White Females Group 1 Face Pool. The antemortem photograph is a black and white candid of average quality scanned from newspaper. The CT slice thickness was 2.5 mm. The subject is edentulous and a wooden dowel was used for articulation of the mandible. Since the antemortem photograph was of the subject at a much younger age, she was selected and included in the younger age demographic face pool.

Subject B (UTK 85-30F)
White female, age 18, weight unknown, height of 65 inches. Photograph 7 in White Females Group 1 Face Pool. The antemortem photograph was a black and white candid of poor quality scanned from a newspaper article. There was full dentition with one slight dental spike extending from a left maxillary molar. The CT slice thickness was 2.5 mm. The subject was selected for a young age demographic (age 18 to 28 years).

Subject C (UTK 17-03D)
White female, age 58, 101 pounds, height of 62 inches. Photograph 5 in White Females Group 2 Face Pool. The antemortem photograph is a color candid photograph cropped around face, and the face is visibly wrinkled. The subject is primarily edentulous and a wooden dowel was used to articulate the mandible. One of the original 30 subjects, the CT slice thickness was 5.0 mm. As ReFace was unable to age subjects, this was one of two subjects selected to test the maximum age range of the prototype (age 50 to 59 years).

Subject D (UTK 35-02D)
White female, age 55, 139 pounds, height of 69 inches. Photograph 1 in White Females Group 2 Face Pool. The antemortem photograph was a color professional portrait of good quality. Wrinkles were visible around eyes and mouth, but less so than Subject C. There was partial dentition with slight dental spikes extending from the skull model. One of the original 30 subjects, the CT slice thickness was 5.0 mm. As ReFace is unable to age subjects, this is one of two subjects selected to test the maximum age range of the prototype (age 50 to 59 years).

Subject E (UTK 81-11F)
White Female, age 25, weight unknown, height of 66 inches. Photograph 5 in White Females Group 1 Face Pool. The antemortem photograph was a color professional portrait of good quality. The CT slice thickness was 2.5 mm. The subject had partial dentition with no metal dental work. This subject was selected for the young age demographic (age 18 to 28 years).

Subject F (UTK 96-10F)
White female, age 28, weight 170 pounds, height of 66 inches. Photograph 1 in White Females Group 1 Face Pool. Antemortem photograph was a poor quality black and white that appears to be from a newspaper. The CT slice thickness was 2.5 mm. The subject had partial dentition with no dental spikes on the skull model. The subject appeared to be younger and thinner than her actual age and weight at death. She was selected and included in the younger age demographic (age 18 to 28 years).

Subject G (UTK 81-09F)
White Male, age 25, 140 pounds, height of 71 inches. Photograph 3 in White Males Group 1 Face Pool. The antemortem photograph was a black and white candid of average quality. The CT slice thickness was 2.5 mm. This subject was selected and included in the optimum age demographic of 25 to 35 years.

Subject H (UTK 02-89D)
White male, age 36, 220 pounds, height of 72 inches. Photograph 5 in White Males

Group 1 Face Pool. The antemortem photograph was an average quality color mug shot of subject at age 32 and of lower weight. The CT slice thickness was 2.5 mm. This subject had partial dentition and there were spikes extending from the skull model. This subject was selected and included in optimum age demographic of 25 to 35 years.

Subject I (UTK 86-18F)
White male, age 26, 145 pounds, height of 70 inches. Photograph 7 in White Males Group 1 Face Pool. The antemortem photograph was a color candid of average quality. The subject had a moustache, but the outline of his face and chin was visible. The CT slice thickness was 2.5 mm. There were slight dental spikes on skull model from dental filling in maxillary molars. This subject was selected and included in the optimum age demographic of 25 to 35 years.

Subject J (UTK 42-05D)
White male, age 42, 175 pounds, height of 74 inches. Photograph 6 in White Males Group 1 Face Pool. The antemortem photograph was a younger color candid of good quality. The subject had a moustache and beard, which obscured the chin outline. The CT slice thickness was 2.5 mm. There was partial mandibular dentition and a maxillary plastic dental appliance with metal attachment clips, but no dental spikes on the skull model. This subject was selected and included in optimum age demographic of 25 to 35 years.

Face Pool Groups
Three different face pools were prepared. Target subjects were divided by sex and age. All target subjects were Americans of European Ancestry. Age at death could not be used exclusively for group selection, as some photographs were obviously taken at a younger age. The exact age in any of the antemortem photographs was unknown. Target subjects were included in the design of the face pool, which asked volunteers to identify any photograph that did not seem to fit with the age range of the others. Subjects with younger photographs than their age at death were consistently identified as not belonging with the group when placed in their actual age categories.

1. White Females (Group 1): Subjects A, B, E, and F and three control photographs. Age range for this group was 18 to 28 years for both target and control photographs. Subject A (age 54) was placed in this group as her antemortem photograph was contemporary with this age group.

2. White Females (Group 2): Subject C (age 58) and Subject D (age 55) and five control photographs. Age range for this group was 50 to 59 years for both target and control photographs.

3. White Males (Group 1): Subjects G, H, I, and J and three control photographs. Age range was 25 to 35 years for both target and control photographs. Subjects H (age 36) and J (age 42) were placed in this group as their antemortem photographs were contemporary with this age group.

Table 8.1 summarizes the target subjects and face pool groupings used in the final validation.

Table 8.1: Demographics for Subjects A through J for Face Pool Validation. Subject B in Resemblance Rating was not 85-30F; a foil approximation and photo that did not match were used as Subject B.

Subjects for Face-Pool Photograhic Validation Tests (N = 10)

UTK ID	11-04D	85-30F	17-03D	35-02D	81-11F	96-10F	81-09F	02-89D	86-18F	42-05D
Test ID	Subject A	Subject B	Subject C	Subject D	Subject E	Subject F	Subject G	Subject H	Subject I	Subject J
Face Pool ID	6	7	5	1	5	1	3	5	7	6
Ancestry	White	White	White	White	White	White	White	White	White	White
Sex	Female	Female	Female	Female	Female	Female	Male	Male	Male	Male
Age	54	18	58	55	25	28	25	36	26	42
Height (inches)	65	65	62	69	66	66	71	72	70	74
Weight (pounds)	120	Unknown	101	139	Unknown	170	140	220	145	175
Face Pool Group	WFG1	WFG1	WFG2	WFG2	WFG1	WFG1	WMG1	WMG1	WMG1	WMG1

Face Pool Group	Number	Code
White Females	Group 1	WFG1
White Females	Group 2	WFG2
White Males	Group 1	WMG1

Participant Groups
The total number of participants for the final validation was 103 adult volunteers. The total group was 67% male and 33% female. None of the participants had expert knowledge; the level of prior knowledge was 0% very, 9% somewhat, and 91% none. The average age for the combined participant groups was 36, with a mode of 28 and a median of 33.5.

Participant Group 1 (PG1): Scientists, included students and staff, N = 21.
Sex: 33% male, 67% female. Level of knowledge: 0% very, 10% somewhat, 90% none.
Age: average 37, mode 23, median 34.

Participant Group 2 (PG2). Experienced state law enforcement officers, N = 21.
Sex: 85% male, 15% female. Level of prior knowledge: 0% very, 15% somewhat, 85% none. Age: average 43, mode 39, median 44.

Participant Group 3 (PG3). Digital photo imaging training class, N = 17.
Sex: 33.33% male, 66.67% female. Level of prior knowledge: 0% very, 11.76% somewhat, 88.24% none. Age: average 38, mode 40, median 45.

Participant Group 4 (PG4). Federal law enforcement trainees, N = 41.
Sex: 86% male, 14% female. Level of prior knowledge (5% somewhat, 95% none).
Age: average 31, mode 26, median 30.

Administration of Validation Tests 1 and 2

Participants were given general instructions prior to the validation testing. A handout was provided to each participant (Appendix A-1). As some of the participants were law enforcement officers and some of the photographs were from actual forensic cases, written and verbal instructions were given that recognition of any photograph would disqualify the participant from the testing. No participant expressed recognition of any target or control photographs during the testings.

In addition to the written instructions, participants were verbally instructed that no envelopes were to be opened until requested to do so by the administrator, envelopes were to be opened one at a time, and answers were not to be discussed with anyone else taking the test. Participants were requested to fill in their sex, age, and level of prior knowledge in the area of forensic facial approximation/reconstruction. Participants were told this information was to be used for statistical analysis only; there was no link between any individual and a specific answer sheet; and that their individual participation and their specific answers were anonymous.

Validation ONE: Face Pool

Each participant was given a sealed envelope with an orange label of "Validation One: Face Pool" taped to the front, and instructions for the Face Pool Validation test were taped to the back (see Appendix A-2). After each participant had received an envelope, the validation test instructions were reviewed and participants were given the opportunity to ask questions about the instructions. Participants were then instructed to open the sealed (outside) envelope and verify receipt of ten sealed envelopes (inside) and an Answer Sheet (see Appendix A-3).

Participants were told that the Answer Sheet was is in alphabetical order from Subject A through Subject J, but the envelopes were not as they had been randomly sorted. Participants were instructed to open the envelopes in order one at a time, make a selection, and mark their answer on the Answer Sheet. After completion of the test, participants were instructed to return all materials to the outside envelope, and to hold the envelope up as an indication of completion to the administrator. Participants were told the administrator would exchange the completed first test for the second test; that the instructions for the second test were the same, but the test was slightly different. As individual participants finished and returned the Validation ONE: Face Pool testing envelope, they were given the second part of the test, Validation TWO: Resemblance Rating.

Validation TWO: Resemblance Rating

Each participant was given a sealed envelope with an green label of “Validation TWO: Resemblance Rating” taped to the front, and instructions for the Resemblance Rating test were taped to the back (see Appendix A-4). As with the prior test, each sealed outside envelope contained ten sealed envelopes and an Answer Sheet (see Appendix A-5).

Time blocks of one or two hours had been established for the entire test, but no specific time limit was given for completion of either test. Participants who took longer to finish the second part of the test were verbally told they could take as much time to complete the test as needed.

Chapter 9
Phase 2: Results of Final ReFace Prototype Validation

Validation Test 1: Face Pool (N = 103)

Three face pools were used to test the ability of humans to correctly identify computerized facial approximations of target subjects with their corresponding antemortem photograph. The results were analyzed with Microsoft® Excel and SPSS 15.0 for Windows, and are expressed in three categories: percents above random chance, Chi square values, and binomial probability. Target hit rates were calculated as the percentage of correct identifications, and random chance was calculated as 1 out of 7 or 14.29%. Percent above random chance was calculated as the hit rate minus 14.29%.

The percent over random chance was used to assess how often the target individual was correctly identified; Chi square values were used to determine whether the frequency distribution for each subject was random or statistically significant. A binomial probability equation was used to assess the correlation between the correct number of answers and the probability those numbers would be obtained in a group of 103 participants by guessing.

<u>Correct Identifications</u>
Nine out of ten target subjects were identified above random chance. Correct target identifications (hit rates) ranged from 6.8% (Subject C) to 60.19% (Subject G). Identification above random chance ranged from -7.49% (Subject C) to 45.9% (Subject G). Subject C had the lowest number of correct identifications and Subject G had the highest number of correct identifications. Figure 9.1 illustrates the overall results for all participant groups.

<u>Chi Squared Results</u>
The Chi Square values indicated the frequency distribution of Face Pool choices was significant for nine out of ten face pool target subjects. (The exception was Subject I, Chi Square 4.311, d *f* 6, P>.05). Figure 9.2 details the frequency distributions and Table 9.1 provides a summary of Chi Square values.

<u>Binomial Probability</u>
The probability that the number of correct identifications *(p)* of the target subjects would be correctly guessed by the total number or participants (*n)* was determined by the binomial probability calculation. The probability (P) was calculated using the general

$$\text{n choose r} \binom{n}{r}$$

which yields the following equation for binomial trials:

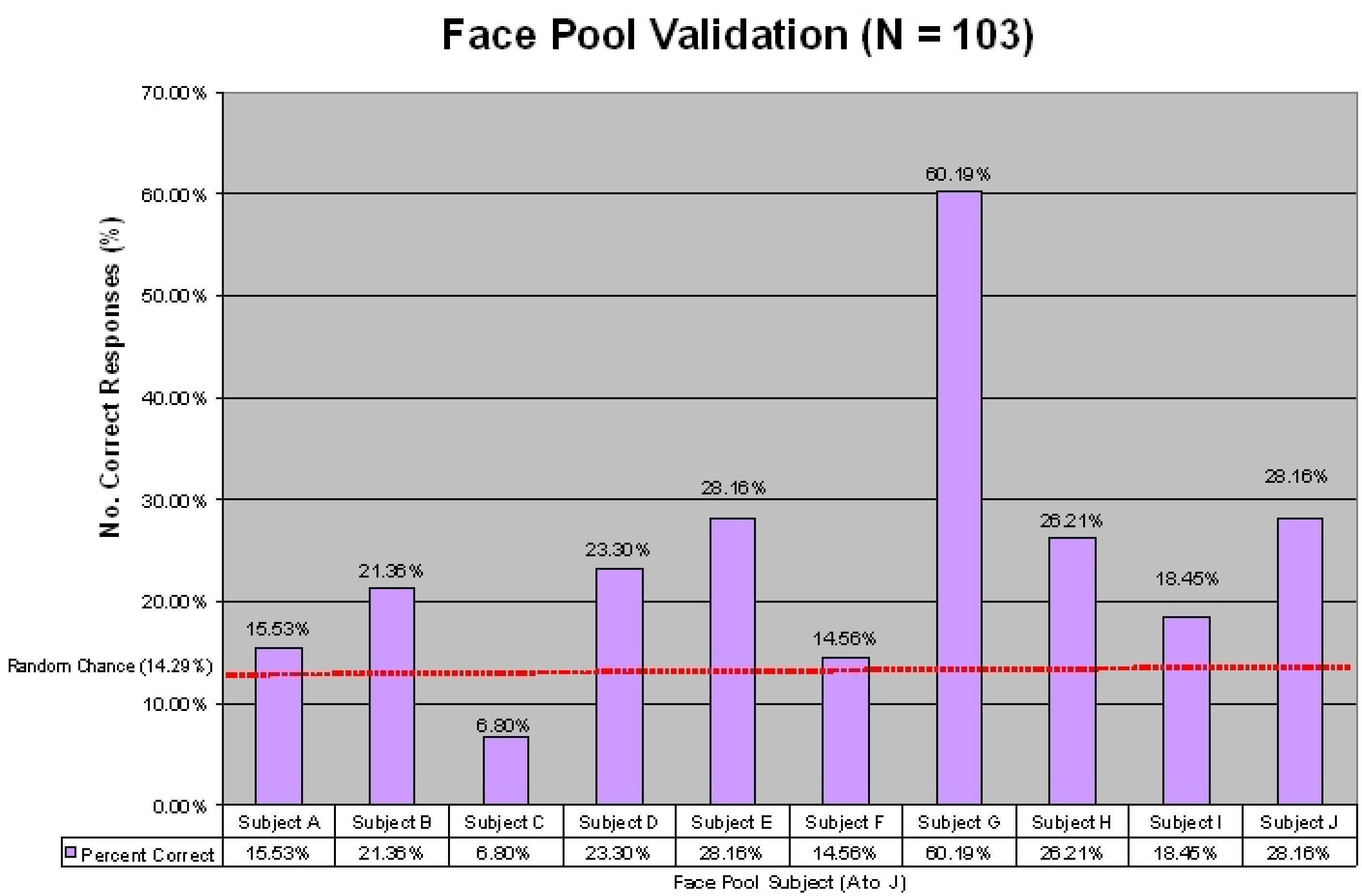


Figure 9.1: Face Pool Validation Results for All Groups (N = 103).

Chi-Square Test

Frequencies

Subject A

Face Pool #	Observed N	Expected N	Residual
1	3	14.7	-11.7
2	16	14.7	1.3
3	26	14.7	11.3
4	25	14.7	10.3
5	15	14.7	0.3
6 (Target)	16	14.7	1.3
7	2	14.7	-12.7
Total	103		

Subject F

Face Pool #	Observed N	Expected N	Residual
1 (Target)	15	14.7	0.3
2	23	14.7	8.3
3	13	14.7	-1.7
4	21	14.7	6.3
5	9	14.7	-5.7
6	21	14.7	6.3
7	1	14.7	-13.7
Total	103		

Subject B

Face Pool #	Observed N	Expected N	Residual
1	18	14.7	3.3
2	7	14.7	-7.7
3	4	14.7	-10.7
4	17	14.7	2.3
5	16	14.7	1.3
6	19	14.7	4.3
7 (Target)	22	14.7	7.3
Total	103		

Subject G

Face Pool #	Observed N	Expected N	Residual
1	4	14.7	-10.7
2	6	14.7	-8.7
3 (Target)	62	14.7	47.3
4	9	14.7	-5.7
5	8	14.7	-6.7
6	8	14.7	-6.7
7	6	14.7	-8.7
Total	103		

Subject C

	Observed N	Expected N	Residual
1	41	14.7	26.3
2	13	14.7	-1.7
3	3	14.7	-11.7
4	2	14.7	-12.7
5 (Target)	7	14.7	-7.7
6	16	14.7	1.3
7	21	14.7	6.3
Total	103		

Subject H

	Observed N	Expected N	Residual
1	17	14.7	2.3
2	11	14.7	-3.7
3	9	14.7	-5.7
4	12	14.7	-2.7
5 (Target)	27	14.7	12.3
6	23	14.7	8.3
7	4	14.7	-10.7
Total	103		

Subject D

	Observed N	Expected N	Residual
1 (Target)	24	14.7	9.3
2	7	14.7	-7.7
3	1	14.7	-13.7
4	6	14.7	-8.7
5	1	14.7	-13.7
6	16	14.7	1.3
7	48	14.7	33.3
Total	103		

Subject I

	Observed N	Expected N	Residual
1	16	14.7	1.3
2	17	14.7	2.3
3	10	14.7	-4.7
4	16	14.7	1.3
5	14	14.7	-0.7
6	11	14.7	-3.7
7 (Target)	19	14.7	4.3
Total	103		

Subject E

	Observed N	Expected N	Residual
1	14	14.7	-0.7
2	7	14.7	-7.7
3	5	14.7	-9.7
4	15	14.7	0.3
5 (Target)	29	14.7	14.3
6	16	14.7	1.3
7	17	14.7	2.3
Total	103		

Subject J

	Observed N	Expected N	Residual
1	13	14.7	-1.7
2	21	14.7	6.3
3	1	14.7	-13.7
4	21	14.7	6.3
5	14	14.7	-0.7
6 (Target)	29	14.7	14.3
7	4	14.7	-10.7
Total	103		

Figure 9.2: SPSS Output for Chi Square Frequencies for Ten Target Subjects.

Table 9.1 Face Pool Validation Summary of Chi Square Values

Test Statistics

	Subject A	Subject B	Subject C	Subject D	Subject E	Subject F	Subject G	Subject H	Subject I	Subject J
Chi-Square[a]	36.388	17.903	74.311	116.039	24.835	25.243	178.427	26.738	4.311	40.058
df	6	6	6	6	6	6	6	6	6	6
Asymp. Sig.	.000	.006	.000	.000	.000	.000	.000	.000	.635	.000

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 14.7.

Chi Squares for six degrees of freedom (Seven face pool subjects (rows) minus 1 = 6*f*)

Critical Values	
10.64	0.10
12.59	0.05
16.81	0.01

$$P(\text{r success out of n trials}) = \binom{n}{r} p^r q^{n-r}$$

where p is the probability of a success and q is the probability of a failure. It follows from the binomial theorem that:

$$\frac{n!}{r!(n-r)!}$$

for the n choose r $\binom{n}{r}$.

The binomial results are expressed as percentages for visual comparison to the actual results obtained by percentages. Table 9.2 provides the binomial probability calculation summary by percent and scientific equation; Table 9.3 provides a comparison between the binomial probability (expressed as percents) to the actual results (expressed by percents).

Binomial Calculation Summary

The binomial probability calculations were used to indicate the probability that the hit rate could have been obtained just by guessing. The binomial calculation results ranged from 6.80E-27 (Subject G) to 11.05% (Subject F), compared with hit rates of 6.8% (Subject C) to 60.19% (Subject G). For Subjects B, D, E, H, and J, the probability was low that the hit rates were likely to be the result of guessing. The probability was higher for Subjects A, C, and F that the results could have been obtained by guessing. Subject I (5.13%) fell between these two groups. The probability calculation indicated that Subject G was the most recognizable.

Results by Face Pool Groupings

White Females (Group 1)

Subjects A, B, C, and three control photographs. Age range for this group was 18 to 28 years for both the target and control photographs. Subject A (age 54) was placed in this group as her antemortem photograph was contemporary with the age group.

Subject A (UTK 11-04D)

Face Pool Photograph 6: Received a hit rate of 16 out of 103 (15.53%), which was 1.24% above random chance. The frequency distribution was significantly above chance expectations (Chi square = 36.389, 6 d_f, $p < .01$). The binomial probability calculation for correctly guessing 16 out of 103 correct guesses is 10.12%, indicating the actual results were too close to guessing to be significant. This was not the most frequently chosen individual for Subject A (Faces 3 and 4 were the most frequently chosen).

Table 9.2: Face Pool Validation: Binomial Probability Calculation Summary.

Subject A	Subject B	Subject C	Subject D	Subject E	Subject F	Subject G	Subject H	Subject I	Subject J
10.12%	1.47%	0.90%	0.48%	0.01%	11.05%	0.00%	0.06%	5.13%	0.01%
1.01E-01	1.47E-02	9.00E-03	4.79E-03	1.17E-04	1.10E-01	6.80E-27	5.99E-04	5.13E-02	1.17E-04

Table 9.3: Comparison of Binomial Probability to Actual Results (%).

Face Pool Validation (N = 103)		
UTK Subject	Binomial P	Actual Result
Subject A	10.12%	15.53
Subject B	1.47%	21.36
Subject C	0.90%	6.80
Subject D	0.48%	23.30
Subject E	0.01%	28.16
Subject F	11.05%	14.56
Subject G	6.80E-27	60.19
Subject H	0.06%	26.21
Subject I	5.13%	18.45
Subject J	0.01%	28.16

Subject B (UTK 85-30F)
Face Pool Photograph 7: Received a hit rate of 22 out of 103 (21.36%), which was 7.07% above random chance. The frequency distribution was significantly above chance expectations (Chi square = 17.903, 6 d_f, $p < .01$). The binomial probability of correctly guessing 22 out of 103 is 1.47%, indicating the result obtained is significant. This was the most frequently chosen individual for Subject B.

Subject E (UTK 81-11F)
Face Pool Photograph 5: Received a hit rate of 29 out of 103 (28.16%), which was 13.87% above random chance. The frequency distribution was significantly above chance expectations (Chi square = 24.835, 6 d_f, $p < .01$). The binomial probability of correctly guessing 29 out of 103 is 0.01%. This was the most frequently chosen individual for Subject E.

Subject F (UTK 96-10F)
Face Pool Photograph 1: Received a hit rate of 15 out of 103 (14.56%), which was 0.27% above random chance. The frequency distribution was significantly above chance expectations (Chi square = 25.243, 6 d_f, $p < .01$). The binomial probability of correctly guessing 15 out of 103 is 11.05%, indicating this result is very close to guessing, This was not the most frequently chosen individual for Subject F. Three other control photographs were chosen more than 20 times each.

White Females (Group 2)

Subject C (age 58) and Subject D (age 55) and five control photographs. Age range for this groups was 50 to 59 years for both target and control subjects.

Subject C (UTK 17-03D)
Face Pool Photograph 5: Received a hit rate of 7 out of 103 (6.80%), which was 7.49% *below* random chance. The frequency distribution was significantly above chance expectations (Chi square = 74.311, 6 d_f, $p < .01$). The binomial probability of correctly guessing only 7 out of 103 is 0.90%. This was the second lowest chosen individual for Subject C. It should be noted another photograph (Face 1) was selected 41 times for this subject.

Subject D (UTK 35-02D)
Face Pool Photograph 1: Received a hit rate of 24 out of 103 (23.30%), which was 9.01% above random chance. The frequency distribution was significantly above chance expectations (Chi square = 116.039, 6 d_f, $p < .01$). The binomial probability of correctly guessing 24 out of 103 is 0.48%. This was not the most chosen individual for Subject D. It should be noted another photograph (Face 7) was selected 48 times for this subject.

White Males (Group 1)

Subjects G, H, I, and J and three foil photographs. Age range was 25 to 35 years for both target and control photographs. Subject H (age 36) and Subject J (age 42) were placed in this group as they appeared younger in their antemortem photographs than age at death. All four target males were the most frequently chosen.

Subject G (UTK 81-09F)
Face Pool Photograph 3: Received a hit rate of 62 out of 103 (60.19%), which was 45.90% above random chance. The frequency distribution was significantly above chance expectations (Chi square = 178.427, 6 d_f, $p < .01$). The binomial probability of correctly guessing 62 out of 103 is an astounding 6.80E-27. This was the most chosen individual in the entire validation.

Subject H (UTK 02-89D)
Face Pool Photograph 5: Received a hit rate of 27 out of 103 (26.21%), which was 11.92% above random chance. The frequency distribution was significantly above chance expectations (Chi square = 26.738, 6 d_f, $p < .01$). The binomial probability of correctly guessing 27 out of 103 is 0.06%. This was the most frequently chosen individual for Subject H, although the distribution for the remaining faces are very close to random chance (14.29%).

Subject I (UTK 86-18F)
Face Pool Photograph 7: Received a hit rate of 19 out of 103 (18.45%), which was 4.16% above random chance. The frequency distribution was *not* significantly above chance expectations (Chi square = 4.311, 6 d_f, $p > .05$). The binomial probability of correctly guessing 19 out of 103 is 5.13%. This was the most frequently chosen individual.

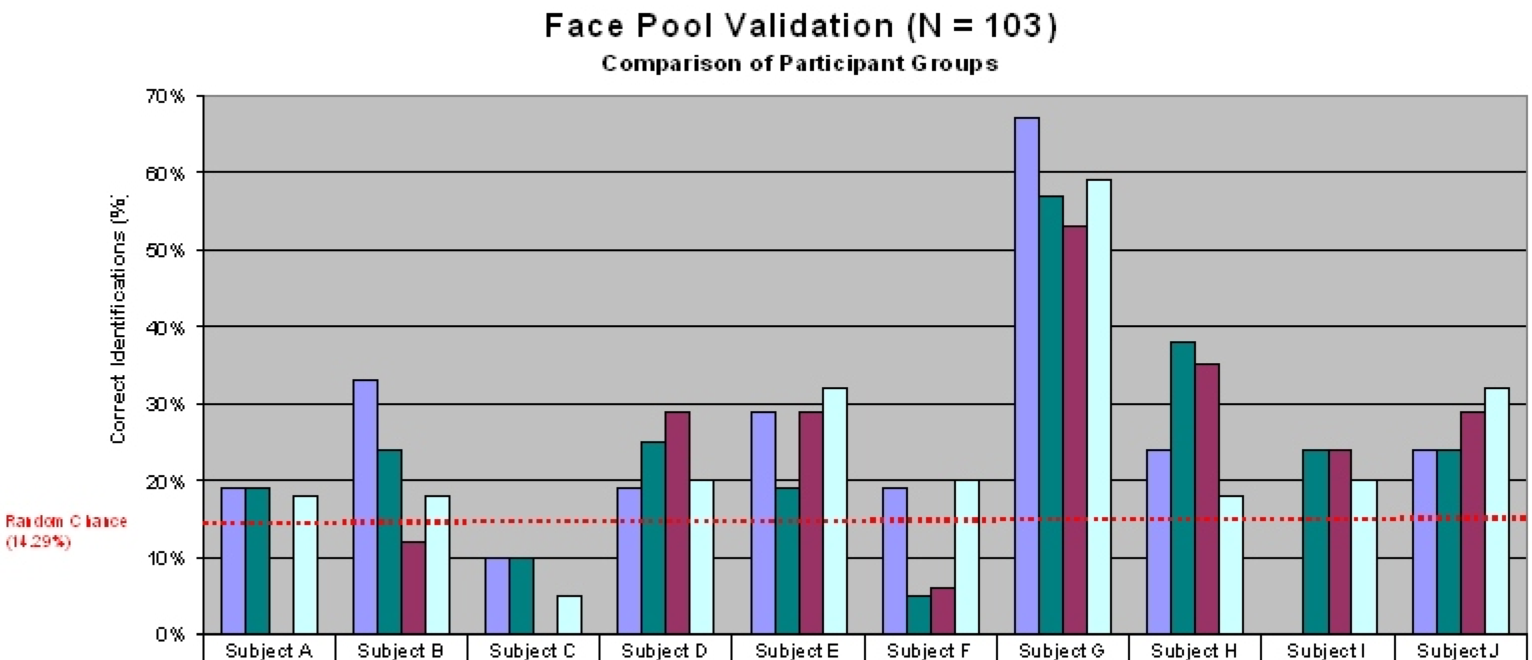
Subject J (UTK 42-05D)
Face Pool Photograph 6: Received a hit rate of 29 out of 103 (28.16%), which was 13.87% above random chance. The frequency distribution was significantly above chance expectations (Chi square = 40.058, 6 d_f, $p < .01$). The binomial probability of correctly guessing 29 out of 103 is 0.01%. This was the most frequently chosen individual for Subject J.

Face Pool Results: Comparison of Participant Groups

Results for each participant group were generally consistent with the overall percentages for combined results (see Figure 9.3). For all groups, Subject G received the highest number of correct identifications, and Subject C was below random chance. Subjects D, E, H, and J were consistent between groups and all were over random chance. Three target subjects received no correct identifications in some groups, reducing the overall results for the total percentages for the combined groups. Subject F was rated below random chance by two groups and above random chance by two groups.

Total number of participants for the final validation was 103 adult volunteers. Sex: 67% male, 33% female. Level or prior knowledge: 0% very, 9% somewhat, 91% none. Age: average 36, mode 28, median 33.5.

Participant Group 1 (PG1): (N = 21). Sex: 33% male, 67% female. Level of knowledge: 0% very, 10% somewhat, 90% none. Age: average 37, mode 23, median 34. *Percent of Correct Identifications*: Subject I received zero correct identifications



	Subject A	Subject B	Subject C	Subject D	Subject E	Subject F	Subject G	Subject H	Subject I	Subject J
PG1	19%	33%	10%	19%	29%	19%	67%	24%	0%	24%
PG2	19%	24%	10%	25%	19%	5%	57%	38%	24%	24%
PG3	0%	12%	0%	29%	29%	6%	53%	35%	24%	29%
PG4	18%	18%	5%	20%	32%	20%	59%	18%	20%	32%

Figure 9.3: Face Pool Validation Comparison of Participant Groups (N = 103).

from this group, and Subject C was the next lowest (10%). Subject B (33%) was higher and Subject G (67%) were both rated higher by this group than any other group. All other percentages of correct identifications were above random chance and consistent with other groups.

Participant Group 2 (PG2): (N = 21). Sex: 85% male, 15% female. Level of prior knowledge: 0% very, 15% somewhat, 85% none. Age: average 43, mode 39, median 44. *Percent of Correct Identifications*: Subject F (5%) was the lowest of any other group, and Subject C (10%) was the second lowest. Subject G (57%) was the highest for this group, and Subject H (38%) was rated higher by this group than any other group. All other percentages of correct identifications were above random chance and consistent with other groups.

Participant Group 3 (PG3): (N = 17). Sex: 33.33% male, 66.67% female. Level of prior knowledge: 0% very, 11.76% somewhat, 88.24% none. Age: average 38, mode 40, median 45. *Percent of Correct Identifications*: Subject A and Subject C received zero correct identifications by this group. Identifications for Subject B (12%) were lower than any other group, and Subject F (6%) was also below random chance. Subject G (53%) was the highest rated for this group, and Subject D (29%) was higher than all other groups. All other percentages of correct identifications were above random chance and consistent with other groups.

Participant Group 4 (PG4). (N = 41). Sex: 86% male, 14% female. Level of prior knowledge (5% somewhat, 95% none). Age: average 31, mode 26, median 30. *Percent of Correct Identifications*: Subject C (5%) was the only subject below random chance for this group. Subject G (59%) was the highest in this group, and Subject E (32%), Subject F (20%), and Subject J (32%) were higher than in any other group. All other percentages of correct identifications were above random chance and consistent with other groups.

Summary of Face Pool Results

One male reconstruction (Subject G) received the majority of correct identifications by all groups. All male target subjects were also the most frequently chosen in the male face pool (White Males Group 1). However, the frequency distribution for Subject I is not statistically significant ($p > .05$), which indicates participants had difficulty with this identification.

Although the frequency distribution was statistically significant, only two of the four female target subjects in the younger age demographic (Subject B and Subject E) received the majority of correct identifications (White Females Group 1). Participants clearly had difficulty with Subject A and Subject F, although it did appear that most were able to eliminate two choices for each face pool, (Subject A, Faces 1 and 7, and for Subject F, Faces 5 and 7).

Subject C and Subject D did not receive the highest number of correct responses in the older demographic face pool (White Females Group 2). The ReFace database contained only seven White females aged 50 to 59, and these two subjects were selected to test the outer age range of the program. If was known that the program could not age subjects, but it was unknown if older subjects could be identified by facial approximations that looked younger. The hit rate for Subject C was the only one *below* random chance. However, a large number of participants who chose the wrong photograph for both subjects chose the *same* control photograph. For Subject C, the same control photograph (Face 1) was selected by 40% of the participants; for Subject D, the same control photograph (Face 7) was selected by 46% of the participants.

Validation Test 2: Resemblance Rating (N = 101)

Nine target subjects from the Face Pool validation were presented in the Resemblance Rating evaluation, and one (Subject B) was replaced by a foil approximation and photograph. The same participants from Face Pool Validation were used for the second validation test. Two participants from Group 2 (PG2) marked more than one answer for a single subject on the Resemblance Rating Answer Sheets. These two tests were removed from the validation, reducing the sample size to N = 101.

Table 9.4 provides a summary of modes and medians for the ten resemblance rating tests. Modes for seven out of ten subjects were close (5) or strong (4) resemblance to the antemortem photograph. The foil comparison, Subject B, received the lowest ratings, with multiple modes of 1 (none) and 2 (slight). The mode for Subject C was 2 (slight), and for Subject E, the mode was 3 (approximate). Table 9.5 provides the frequency of resemblance ratings by subject. Figure 9.4 illustrates the rating distribution by number of responses, and Figure 9.5 illustrates the rating responses by percents.

For comparison purposes, rating categories of none and slight were combined, and rating categories of close and strong were combined, (see Figure 9.6). The category of approximate was not combined with any other category. Over 60% of the participants rated the foil (Subject B) as none or slight resemblance. Consistent with the Face Pool results, Subject C received the lowest ratings (75% none or slight), and Subject G received the highest ratings (65% close or strong).

ReFace approximations for Subject A (48%), Subject D (54%), Subject E (54%), Subject G (65%), Subject H (55%), and Subject J (40%) were all rated as a close or strong resemblance to their antemortem photographs. On the other end of the rating scale, Subject C (75%) received even lower ratings that the foil comparison, Subject B (61%). Subject I and Subject F received similar percentages in both the low and high categories.

Resemblance Ratings by Face Pool Groupings

White Females (Group 1)

<u>Subject A (UTK 11-04D)</u>
The highest category was 4 (close), with a frequency of 29. The mode was also 4 and the median was 3 (approximate). The combined percentage for categories close and strong was 48%.

Table 9.4: Resemblance Rating Validation Test (Modes and Medians)

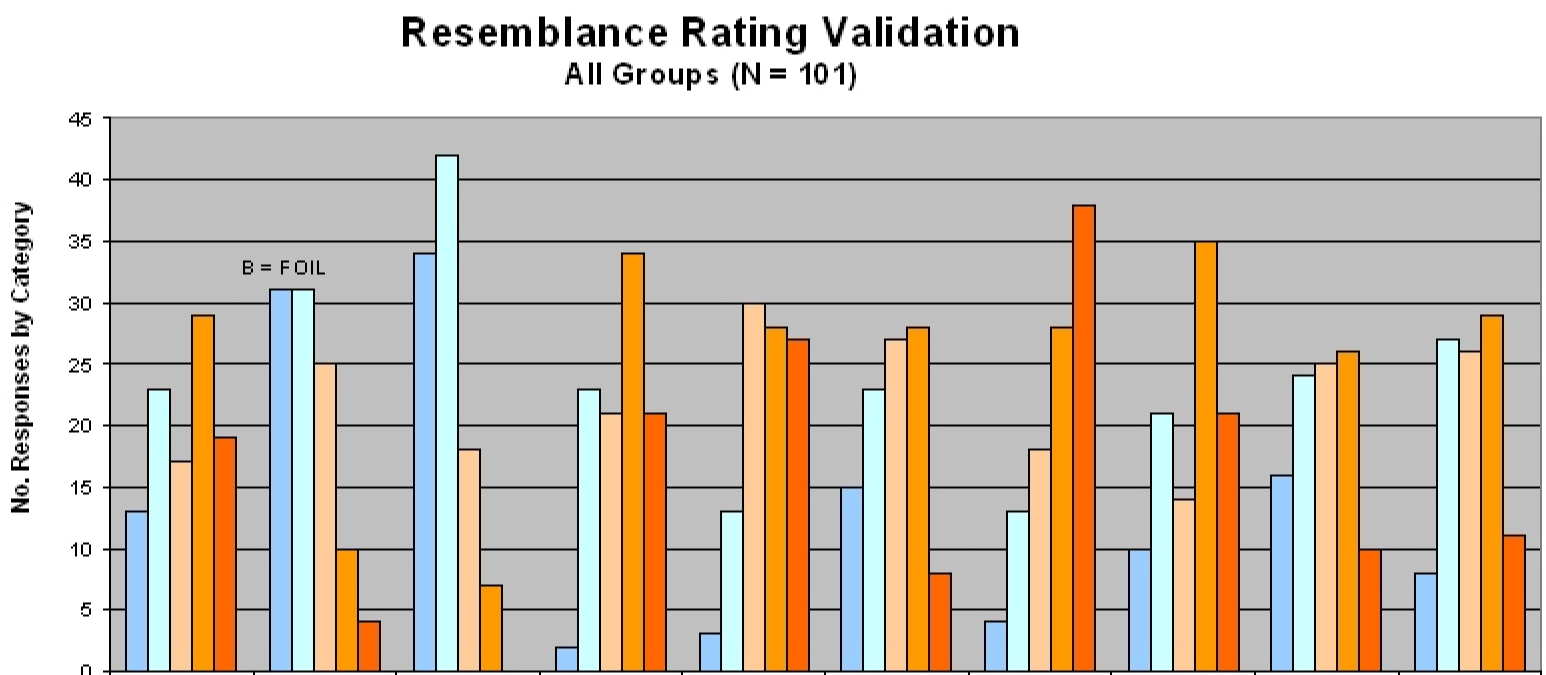
Statistics

		SubjectA	Subject B	Subject C	Subject D	Subject E	Subject F	Subject G	Subject H	Subject I	Subject J
N	Valid	101	101	101	101	101	101	101	101	101	101
	Missing	0	0	0	0	0	0	0	0	0	0
Median		3.00	2.00	2.00	4.00	4.00	3.00	4.00	4.00	3.00	3.00
Mode		4	1[a]	2	4	3	4	5	4	4	4

a. Multiple modes exist. The smallest value is shown

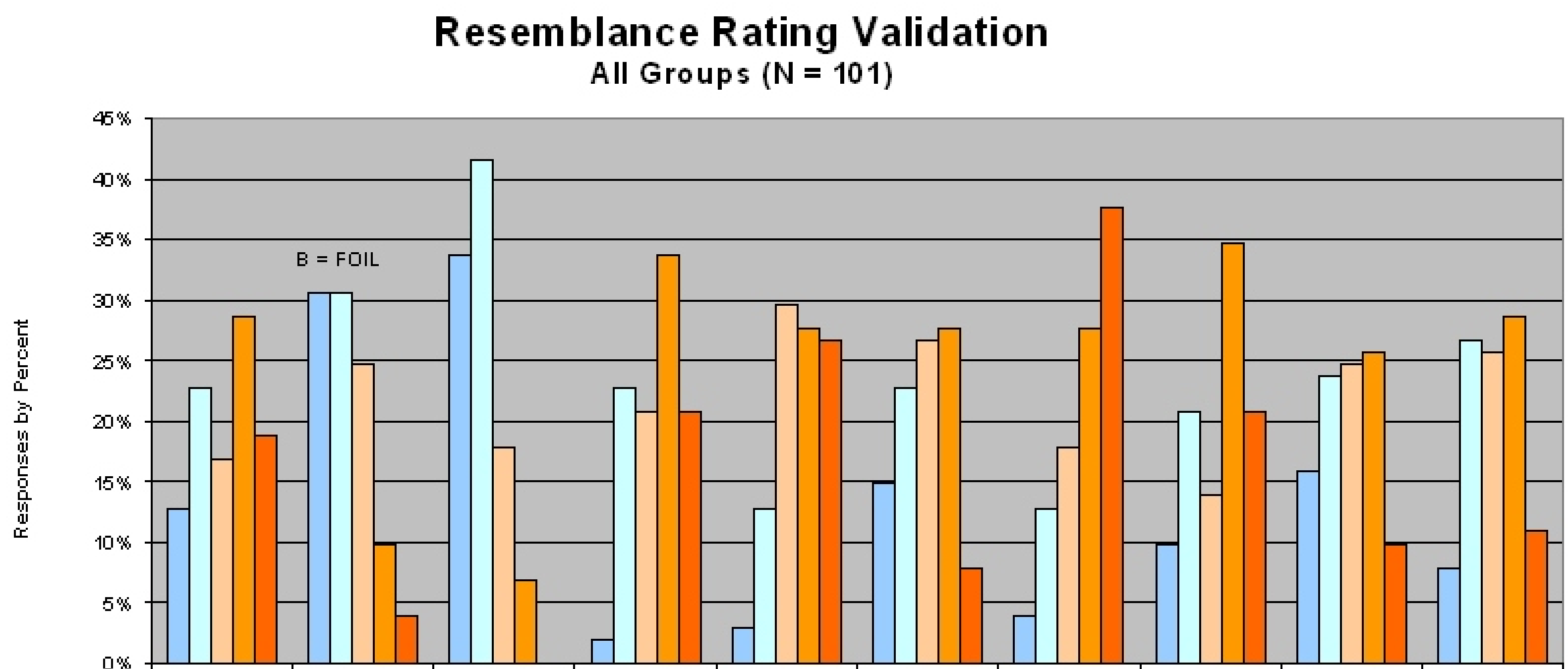
Table 9.5: Frequency of Resemblance Ratings by Subject (N = 101).

Resemblance Rating Scale	Subject A Count	Subject B Count	Subject C Count	Subject D Count	Subject E Count	Subject F Count	Subject G Count	Subject H Count	Subject I Count	Subject J Count
1 = None	13	31	34	2	3	15	4	10	16	8
2 =Slight	23	31	42	23	13	23	13	21	24	27
3 = Approximate	17	25	18	21	30	27	18	14	25	26
4 = Close	29	10	7	34	28	28	28	35	26	29
5 = Strong	19	4	0	21	27	8	38	21	10	11



	Subject A	Subject B	Subject C	Subject D	Subject E	Subject F	Subject G	Subject H	Subject I	Subject J
1 = None	13	31	34	2	3	15	4	10	16	8
2 = Slight	23	31	42	23	13	23	13	21	24	27
3 = Approximate	17	25	18	21	30	27	18	14	25	26
4 = Close	29	10	7	34	28	28	28	35	26	29
5 = Strong	19	4	0	21	27	8	38	21	10	11

Figure 9.4: Resemblance Ratings Responses by Number of Responses for All Participant Groups (N = 101).



	Subject A	Subject B	Subject C	Subject D	Subject E	Subject F	Subject G	Subject H	Subject I	Subject J
1 = None	12.87%	30.69%	33.66%	1.98%	2.97%	14.85%	3.96%	9.90%	15.84%	7.92%
2 = Slight	22.77%	30.69%	41.58%	22.77%	12.87%	22.77%	12.87%	20.79%	23.76%	26.73%
3 = Approximate	16.83%	24.75%	17.82%	20.79%	29.70%	26.73%	17.82%	13.86%	24.75%	25.74%
4 = Close	28.71%	9.90%	6.93%	33.66%	27.72%	27.72%	27.72%	34.65%	25.74%	28.71%
5 = Strong	18.81%	3.96%	0.00%	20.79%	26.73%	7.92%	37.62%	20.79%	9.90%	10.89%

Figure 9.5: Resemblance Rating Responses by Percents for All Participant Groups (N = 101).

Table 9.6: Frequency Distribution with Combined Categories (none and slight, and close and strong).

Rating Scale	Frequency Distribution by Percents (N = 101)									
	Subject A	Subject B	Subject C	Subject D	Subject E	Subject F	Subject G	Subject H	Subject I	Subject J
1-2	36%	61%	75%	25%	16%	38%	17%	31%	40%	35%
3	17%	25%	18%	21%	30%	27%	18%	14%	25%	26%
4-5	48%	14%	7%	54%	54%	36%	65%	55%	36%	40%
	N = 101	N = 101	N = 101	N = 101	N = 101	N = 101	N = 101	N =101	N = 101	N = 101
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Resemblance Rating Scale

1-2	None or Slight
3	Approximate
4-5	Close or strong

Subject B (UTK 85-30F)
Subject B was the foil in the Resemblance Rating test: the facial approximation and the antemortem photograph were of different individuals. Neither were the actual Subject B from the Face Pool test. The highest frequencies were tied between 1 (none) and 2 (slight), each selected 31 times. The mode was tied; the smallest value was 1 (none), and the median was 2 (slight). The combined percentage for categories none and slight was 61%.

Subject E (UTK 81-11F)
The highest category was 3 (approximate), with a frequency of 30. The mode was 3 and the median was 3. However, the frequencies for both close (28) and strong (27), were similar. The combined percentage for categories close and strong was 54%.

Subject F (UTK 96-10F)
The highest category was 4 (close), with a frequency of 28; the frequency for 3 (approximate) was 27. The mode was 3 with a median of 4. The combined percentage for categories close and strong was 36%. Notable is the combined percentage for categories slight and none (38%).

White Females (Group 2)

Subject C (UTK 17-03D)
The highest category was 2 (slight), with a frequency of 42. Both the mode and median were also 2. The combined percentage for categories none and slight was 75%.

Subject D (UTK 35-02D)
The highest category was 4 (close), with a frequency of 34. Both the mode and median were also 4. The combined percentage for categories close and strong was 54%.

White Males (Group 1)

Subject G (UTK 81-09F)
The highest category was 5 (strong), with a frequency of 38. This was the only subject with the highest frequency in category 5. The mode was 5 and the median was 4. The combined percentage for categories close and strong was 65%.

Subject H (UTK 02-89D)
The highest category was 4 (close), with a frequency of 35. Both the mode and the median were also 4. The combined percentage for categories close and strong was 65%.

Subject I (UTK 86-18F)
The highest category was 4 (close), with a frequency of 26. However, the frequencies for both slight (24) and approximate (25) were similar. The mode was 3 and the median was 4. The combined percentage for categories close and strong was 36%. The was the only subject in the Face Pool whose Chi square score was not significantly above random chance. The combined percentage for categories close and strong was 36%. The combined categories for none and slight was 40%.

Subject J (UTK 42-05D)
The highest category was 4 (close), with a frequency of 29. The mode was 4 and the median was 3. The combined percentage for categories close and strong was 40%. The combined categories for none and slight was 35%.

Resemblance Ratings: Comparison of Participant Groups

Results for each participant group were generally consistent with the overall percentages for combined results, (see Figure 9.7). For all groups, Subject G received the highest resemblance ratings. In three out of four groups, resemblance ratings for Subject C were even lower than those of the foil comparison, Subject B. Subjects D, E, H, and J were rated as close or strong resemblance by a majority of the participants. Resemblance ratings for Subjects F, I, and J have the most group variation.

Group statistics for the Resemblance Rating test vary slightly from the Face Pool test, and the revised statistics are listed below. Variation was because some participants changed their level of expertise from the first to the second test, two answer sheets were incomplete and removed from the validation, and different participants left the age, sex, or level of knowledge categories blank.

Participant Group 1 (PG1): (N = 21). Sex: 33% male, 67% female. Level of knowledge: 5% very, 10% somewhat, 86% none. Age: average 37, mode 26, median 34. *Resemblance Ratings*: PG1 was the only group that rated the foil, Subject B (76%), higher than Subject C (52%) in the combined categories of none or slight resemblance. Subject F was rated as close or strong resemblance by this group at a higher percentage (52%) than any other group.

Participant Group 2 (PG2): (N = 19). Sex: 89% male, 11% female. Level of prior knowledge: 0% very, 17% somewhat, 83% none. Age: average 43, mode 39, median 44.5. *Resemblance Ratings*: PG2 was the only group that rated Subject I as the highest percentage in the close and strong combined categories (53%). Subject F (26%) and Subject I (21%) were rated lower in the close and strong combined categories than any other group. Subject H was rated highest in the close and strong combined categories (68%) than in any other group.

Participant Group 3 (PG3): (N = 17). Sex: 33% male, 67% female. Level of prior knowledge: 0% very, 12% somewhat, 88% none. Age: average 38, mode 40, median 35. *Resemblance Ratings:* PG3 was the only group to rate Subject A highest in the combined none or slight categories (47%). Percentages of responses for Subjects F, I, and J were equal in both combined categories (none and slight, close and strong).

Participant Group 4 (PG4). (N = 44). Sex: 86% male, 14% female. Level of prior knowledge (5% somewhat, 95% none). Age: average 31, mode 26, median 30. *Resemblance Ratings:* Subject I was rated lowest in the close and strong combined categories than in any other group (27%). This was the largest test group, and the percentages obtained from this group are consistent with the overall percentages.

Rating Scale	Resemblance Rating: GROUP 1: Frequency Distribution by Percents (N = 21)									
	Subject A	Subject B	Subject C	Subject D	Subject E	Subject F	Subject G	Subject H	Subject I	Subject J
1-2	29%	76%	52%	24%	5%	14%	14%	29%	43%	43%
3	24%	14%	24%	10%	43%	33%	5%	19%	24%	14%
4-5	48%	10%	24%	67%	52%	52%	81%	52%	33%	43%
	N=21	N=21	N=21	N=21	N=21	N=21	N=21	N=21	N=21	N=21
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Rating Scale	Resemblance Ratings: GROUP 2: Frequency Distribution by Percents (N = 19)									
	Subject A	Subject B	Subject C	Subject D	Subject E	Subject F	Subject G	Subject H	Subject I	Subject J
1-2	32%	58%	79%	16%	26%	53%	16%	21%	21%	42%
3	11%	21%	21%	32%	11%	21%	26%	11%	26%	37%
4-5	58%	21%	0%	53%	63%	26%	58%	68%	53%	21%
	N=19	N=19	N=19	N=19	N=19	N=19	N=19	N=19	N=19	N=19
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Rating Scale	Resemblance Ratings: GROUP 3: Frequency Distribution by Percents (N = 17)									
	Subject A	Subject B	Subject C	Subject D	Subject E	Subject F	Subject G	Subject H	Subject I	Subject J
1-2	47%	53%	88%	24%	12%	41%	24%	24%	41%	47%
3	18%	24%	12%	12%	29%	18%	18%	29%	18%	6%
4-5	35%	24%	0%	65%	59%	41%	59%	47%	41%	47%
	N=17	N=17	N=17	N=17	N=17	N=17	N=17	N=17	N=17	N=17
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Rating Scale	Resemblance Ratings: GROUP 4: Frequency Distribution by Percents (N = 44)									
	Subject A	Subject B	Subject C	Subject D	Subject E	Subject F	Subject G	Subject H	Subject I	Subject J
1-2	36%	59%	80%	30%	18%	41%	16%	39%	45%	23%
3	16%	32%	16%	25%	32%	30%	20%	7%	27%	34%
4-5	48%	9%	5%	45%	50%	30%	64%	55%	27%	43%
	N=44	N=44	N=44	N=44	N=44	N=44	N=44	N=44	N=44	N=44
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Figure 9.6: Resemblance Ratings by Percent by Participant Groups. Categories of None or Slight (1-2), and Close or Strong (4-5) are Grouped, and percents rounded.

Summary of Face Pool and Resemblance Rating Results

Subject C received the lowest hit rate and resemblance rating scores, and Subject G received the highest hit rate and resemblance rating scores. There were no significant differences in the hit rates between male participants (67%) and females participants (33%) or between participant groups. All participants were non-experts.

The hit rate in the Face Pool test was above random chance for nine out of ten target subjects. The exception was Subject C, whose hit rate was less than random change (6.8%). The range was from 6.80% to 60.19%, with a mean hit rate of 24%, or 10% above random chance. Male approximations received a higher frequency of correct identifications than female approximations.

Overall resemblance rating scores were 13% none, 24% slight, 22% approximate, 25% close, and 16% strong. Subjects received the highest number of ratings for close resemblance (six subjects), strong resemblance (Subject G), approximate resemblance (Subject E), and slight resemblance (Subject C). The foil subject received equal numbers of ratings for no resemblance (30.5%) and slight resemblance (30.5%).

Discussion

Comparison with Previous Research

Few previous studies were found that compared antemortem photographs with facial approximations using face pools or resemblance ratings. This research validation project appears to be the first to use such methods to test *computer* generated facial approximations with antemortem photographs, so the results of this study cannot be directly compared with previous research.

A method of standardizing results from different research studies using facial approximations and face pools is to compare the percentage of correct responses above random chance. For this research, random chance was defined for the face pool as 1 out of 7, or 14.29%. The variation in percent above random change between subjects ranged from 7.29% *below* random chance to 45.90% *above* random chance. The variation in previous studies ranged from 19% below random chance (Stephen and Henneberg 2001) to 88% above random chance (Stephan and Arthur 2006). However, some researchers did not specify results as hit rates or percents over random chance and these rates must be extrapolated from the research by those who wish to compare them to other studies. Such overall results do not reflect the extreme variation between individual target subjects, and the variation in research design between researchers may have also influenced the success or failure rates.

<u>Variation in Research Design</u>

Research design and results were compared for the following studies: Snow et al 1970, Helmer et al 1993, Stephan and Henneberg 2001, Wilkinson and Whittaker 2002, Stephan and Arthur 2006, and Wilkinson et al 2006. The number of target subjects ranged from one to ten. The number of approximations prepared from a single target subject ranged from two to sixteen. The number of photographs used in the face pools ranged from five to ten. The number of individuals who did the comparisons ranged from 14 to 200. The percent of correct identifications ranged from 2% to 98%.

Variation in Age of Target Subject and Age in Facial Approximation

Extreme variation in the range of results was found in the research study of Snow and colleagues (1970). Two skulls were used (one male and one female) for two facial approximations (one of each subject). Each approximation was presented in a face pool of seven photographs (the target and six controls). The female approximation was rated by 104 individuals; the male approximation was rated by 200 individuals. The results of the female approximation were 26% correct or 11% over chance; the results for the male approximation were 67% correct or 54% above chance. The mean hit rate of 33% does not reflect the extreme variation between two subjects whose approximations were prepared by the same experienced artist (Betty Pat Gatliff).

The ReFace validation also reflected extreme variation between a younger male (Subject G, age 25), and an older female (Subject C, age 58). Both approximations were automatically generated by the same process. Each approximation was presented in a face pool of seven photographs (the target and six controls). Both approximations were rated by 103 individuals. The inability of the program to age the facial approximation for Subject C meant the facial approximation appeared to be much younger than her actual age. The facial approximation for Subject G was consistent with his age. The results of the female approximation were 6.8% correct or 7.49% *below* random chance; the results of the male approximation were 45.90% above random chance.

Variation in Comparison Images

Of obvious difference in the present study is that the computerized facial approximations are without hair, have closed eyes, no skin texture, and are uniform in color. Similar looking images, called “head models,” were used by Bruce and colleagues in an experimental study (1991) to test the ability to match unfamiliar faces using computerized images and photographs. The head models were created by skin surface laser scans, and had no pigmentation and minimal skin texture. Results of this study indicated that female head models were disproportionately more difficult to recognize than male head models. (The results of the ReFace validation were consistent in that both sexes were better able to identify male target subjects). Photographs of individuals with closed eyes and covered hair were better identified than the head models, indicating the difficulty in identification was not solely due to the lack of hair and closed eyes. The addition of hair to the head models did not improve accuracy (Bruce and Young 1998, Bruce et al 1991).

Wilkinson and colleagues (2006), compared computerized facial approximation images with face pool images derived from conversion of CT DICOM data into stereolithography (STL) files of skeletal and skin surface images. Both the facial approximation target subjects and the face pool subjects were without skin texture or hair, with closed eyes, and of uniform color. In this comparison of similar images, the hit rate was 70% (50% above chance). The high success rate of this accuracy test may be that the images being compared were similar in appearance, viewpoint (pose), and lighting.

Matching similar facial images is less difficult than if there are variations in the images being compared (Hill and Bruce 1996, Bruce et al 1999). As actual antemortem photographs were used for the target and control photographs, the color, scale, and resolution varied. Attempts were made to standardize the images, but there was obvious variation between the images. This use of antemortem photographs would better reflect a forensic setting in which a facial approximation might be compared to photographs of missing persons. However, the dissimilarity between images may contribute to less accurate results in a research environment.

Expert vs Non-Expert

Previous research has indicated that experts score higher than non-experts in face pool validations. None of the participants in the final ReFace validation were experts. They were not told that soft tissue features such as the ears, neck, and tip of the nose were generic features. Photographs may have been discounted due to variation between the photographs and these generic features, whereas experts would have known to discount them in the comparisons. This was confirmed by some participants after the validation test was completed, who indicated it took several comparisons before realization that the ears and neck were generic. Differences in the location of the nose tip was frequently mentioned as a method of elimination of possible target subjects from the face pool.

Male versus Female

In this research study, the ages of male target subjects were within the optimum age range of the ReFace database, whereas the female target subjects were at either the low or high range of database subject age. The higher rate of identification by both sexes for the male target subjects may have been due to the better representative age group sample from the database. In a face pool of unfamiliar photographs (Snow et al 1970) and in a psychological study in the identification of familiar photographs (Doty 1998), greater accuracy by both sexes was found in the identification of male target subjects.

Visible Wrinkles and Youthful Approximations

Two target subjects (Subject C and Subject D) and five control photographs were used for the face pool, Group 2: White Females). These two subjects were included as a test of the outer age range of the program. It was unknown if participants could match photographs of target subjects that appeared older than the facial approximation. The below random chance results for Subject C would seem to indicate that they could not. However, the results for Subject D were consistent with the younger target subjects. Both subjects were of similar age, yet the antemortem photograph for Subject C showed visible wrinkles around the eyes and mouth, whereas Subject D had slight wrinkles around the eyes and mouth. These results would seem to indicate that it was the frequency of wrinkles, and not the variation in age, that caused the low number of correct identification for Subject C.

Design of Face Pool

The majority of participants did not correctly identify Subject C or Subject D; however, they did choose the same incorrect photograph. For Subject C, 40% of participants chose Face 1 in the face pool (who was actually Subject D). For subject D, 47% chose Face 7, a control photograph. Subject C had many visible wrinkles; Subject D had wrinkles primarily around the eyes. The control photograph in Face 7 was of the same age range, but wrinkles were not apparent due to the quality of the photograph.

Face 1 and Face 7 were similar in face shape and appearance, and each received more than 60% of the total responses in the face pool Group 2: White Females. This may indicate the photographs were too similar and should not have been included in the same face pool. The variation between the two faces was in the quality and resolution of the antemortem photographs and in the visible appearance of wrinkles on the face. The antemortem photograph for Subject D was a professional portrait, with wrinkles visible around the eyes. Face 7 was cropped from a candid photograph, with poor resolution and no obvious appearance of wrinkles.

The below random chance rating for Subject C was an anomaly in the validation test. The poor result may have been influenced by one or more of the following: inclusion of similar photographs in the face pool, variation in the quality and resolution of the photographs most frequently selected, variation in the number and frequency of face wrinkles, and younger appearing facial approximations.

Comparison with Face Perception and Recognition Research

Familiar versus Unfamiliar Faces

A research study by Megreya and colleagues (2006) found that is was difficult to match two images of the same *unfamilia*r face, and concluded the identification process for familiar and unfamiliar faces are processed in qualitatively different ways. As applied to facial approximations, research testing uses unfamiliar target and control subjects for matching, whereas forensic applications would be attempting to have a match made by someone familiar with the subject of the approximation. Thus, high hit rates in research testing may not be indicative of success in forensic settings.

Facial approximations used in the Phase 2 final validation had closed eyes and were without facial or head hair. In a study by Doty (1998), recognition of familiar photographs indicated the facial characteristic identified as most useful in face identifications for "Caucasians" was hair, whereas other ancestral groups chose face shape, eyes, lips, and nose (but never hair). However, an experimental study by Bruce and colleagues (1991) found that the addition of hair to computer generated faces did not improve accuracy in matching of unfamiliar images.

Presentation of Images

ReFace facial approximations can be viewed and rotated in three-dimensional virtual space; however the face pool images were presented as 2D photographs. The kinetic depth effect is a perceptual phenomenon whereby rotating a 2D image transforms the it into a 3D image (Wallach and O'Connell 1953). Presentation of a rotating facial approximation to groups or to individuals may improve the identification of target subjects in forensic settings, but may not improve the accuracy in research settings. Bruce and colleagues (1999) attempted to ascertain whether accuracy would be improved by showing targets in video clips rather than photographs. They found consistent performance for both video and photographs. Target subjects that were difficult to identify in one medium were also difficult to identify in the other.

Hill and Bruce (1996) found that variations in lighting increased the difficulty in matching faces, even if the viewpoint (pose) was the same. The facial approximations were presented in front, profile, and three-quarter views. The viewpoint of the antemortem photographs varied from full face to three-quarter, and most had some degree of horizontal or vertical tilting of the face. Lighting for the facial approximations was consistent; however the lighting for the target and control photographs was highly variable. For research testing, the use of antemortem photographs for comparison of facial approximations can clearly indicate whether or not the facial approximation is an accurate representation of the living person. However, in research, the inability to control the lighting and viewpoint in such photographs may decrease the accuracy of target identifications.

Nationality and Ancestry of Participants and Target Subjects
All target subjects, and most of the participants, were Americans of European ancestry (White). The participants were requested to indicate their sex, age, and level of prior knowledge, but not ancestry or nationality. Doty (1998) found that individuals were more accurate in the identification of previously witnessed faces when the target subjects were of the same nationality, and extrapolated from this finding that eyewitness identification of a person from a different nationality would be less reliable. Based on the differences found in the identification of familiar versus unfamiliar faces, this research may have some bearing if the facial approximation targets and the participants attempting to identify them are of different nationalities.

Conclusions

The results of the validation study seem to confirm that humans can match antemortem photographs of target subjects with computerized facial approximations. Results for the Face Pool Validation and the Resemblance Rating Validation indicate that for the majority of target subjects, the identifications were neither random nor guesses.

The design, development, and validation of any prototype is often painfully slow, with incremental advances in performance and accuracy. ReFace has made continued progress toward a fully functioning computerized facial approximation program. Beta Version 7.0 contains automated methods of increasing age and adjusting weight to the program default of average. Eyes can be opened to give the approximation a more human appearance. The addition of facial and head hair is in development.

Currently, the ReFace database is population specific, as are traditional tissue depth measurements. The database subjects are all Americans with self-identified European, African, or Asian *ancestry*; none are natives of Europe, Africa, or Asia. However, any population can be added to the database. A statistically representative sample can be formed with CT scans of 100 living members of a population (50 males and 50 females) who ideally represent a representative sampling of ages.

The development focus has been on the forensic identification on unknown human *skeletal* remains (as are traditional methods). However, ReFace can build a facial approximation from a CT scan of *decomposing* human remains. Removal of the head from the body or defleshing of the skull are not required.

Future Research and Development

Faster and more cost effective facial approximations based on better data do not guarantee better results. Two areas that have not been resolved are subjectively and results based on population averages of soft tissue depths. Neither traditional nor computer facial approximations can accurately predict the phenotypic expression of facial soft tissues. The variation of the position of nose tip or ear size and position cannot be accurately ascertained by current methods. Indicators of aging, such as jowls and wrinkles, do not lend well to population averages as they have both genetic and environmental components.

New technology has provided updated tissue depth measurements on larger populations of different nationalities (see Manhein et al 2000 for Americans, and DeGreef et al 2006 for Europeans). Differences in tissue depths obtained from such studies may reflect variation in methods of obtaining measurements, or may be confirming that such measurements are population specific. Traditionally classified European “Caucasians” may or may not have the same tissue depths as American “Caucasians.”

Traditional and computerized methods of facial approximation both rely on some method of averaging to determine the specific depths for a specific facial approximation. Since the exact size and shape of facial soft tissue features is unknown, an average is still the best we can do. Such averages may hinder identification of an individual who has very distinctive soft tissue features such as ears that stick out, a large nose, or uneven skin tones, acne, or facial scars. For now, averages are the best that can be done. However, a facial approximation with soft tissue features based on population averages could hinder identification of a person with extreme variation from average.

There has never been a known identification based solely on a computerized facial approximation. As the ReFace prototype demonstrates, such images may be available for forensic cases in the near future. Important future issues to be addressed are the determination of the presentation method and the intended audience. Traditional facial approximations have been shown to the general public in newspapers, television newspapers, and web sites. Whether this would be the best presentation of computerized images needs to be determined. Whether the public would perceive the computerized image as equivalent to a traditional artistic image is unknown. The ability for rotation of a 3D face may increase the possibility of identification. The best method of presentation of the images depends on who will see them. Possibilities are other computers with facial recognition software; forensic artists, experts who can add more realistic skin texture and hair; law enforcement officers who may or may not have artistic training; or the general public through mass media.

No one knows the exact number of unidentified human remains in the United States. There is no agency, group, or government entity responsible for compiling, maintaining, and reporting this information. There are no federally mandated requirements for any jurisdiction in the United States to report this information. There are an estimated 100,000 active missing persons cases in the United States on any given day. Hundreds of thousands of people have simply disappeared. There is no centralized database to search for missing persons.

The Bureau of Justice Statistics is currently conducting a census to ascertain how many unidentified sets of human remains are actually being stored by medical examiners and coroners in the United States. Preliminary information from this census is that there are more than 40,000 unidentified sets of human remains being stored right now. However, the FBI's National Crime Information Center (NCIC) database is estimated to contain only 10% to 15% of the total number. As of May 8, 2007, there are 6,246 reported cases of unidentified deceased individuals in the NCIC database. Reporting cases to the NCIC is voluntary. The majority of records come from a single state, California, as it is the only state that requires reporting of cases to the NCIC.

Our crisis of missing persons and unidentified remains has been called our Nation's silent mass disaster (Ritter 2007). ReFace is in the final phase of testing, and may soon be able to provide an accurate, rapid, and cost effective method to assist in the identification of unidentified victims.

List of References

List of References

Aulsebrook, W. A., P. Becker, and M. Y. Iscan. 1996. Facial Soft-Tissue Thickness in the Adult Male Zulu. *Forensic Science International* 79:83-102.

Aulsebrook, W. A., M. Y. Iscan, J. H. Slabbert, and P. Becker. 1995. Superimposition and reconstruction in forensic facial identification: A survey. *Forensic Science International* 75:101-120.

Bartlett, J. C., S. Hurrey, and W. Thorley. 1984. Typicality and Familiarity of Faces. *Memory and Cognition* 12:219-228.

Brown, R. E., T. P. Kelliher, P. H. Tu, W. D. Turner, M. A. Taister, and K. W. P. Miller. 2004. A Survey of Tissue-Depth Landmarks for Facial Approximation. *Forensic Science Communications* 6:1-27.

Bruce, V. 1982. Changing Faces: Visual and Non-visual Coding Processes in Face Recognition. *British Journal of Psychology* 73:105-116.

Bruce, V. 1986. Influences of Familiarity on the Processing of Faces. *Perception* 15:387-397.

Bruce, V. 1988. *Recognising Faces.* Hillsdale, NJ: Laurence Erlbaum.

Bruce, V., T. Doyle, A. Coombes, P. Healey, M. Burton, and A. Linney. 1991. Recognising Facial Surfaces. *Perception* 20:755-69.

Bruce, V., Z. Henderson, K. Greenwood, P. J. B. Hancock, A. M. Burton, and P. Miller. 1999. Verification of face identities from images captured on video. *Journal of Experimental Psychology-Applied* 5:339-360.

Bruce, V., T. Valentine, and A. Baddeley. 1987. The Basis of the 3/4 View Advantage in Face Recognition. *Applied Cognitive Psychology* 1:109-120.

Bruce, V., and A. Young. 1998. *In the Eye of the Beholder: The Science of Face Perception*, Paperback 2000 edition. New York: Oxford University Press Inc.

Burton, A. M., R. Jenkins, P. J. B. Hancock, and D. White. 2005. Robust representations for face recognition: The power of averages. Cognitive Psychology 51:256-284.

Byers, S. N. 2002. *Introduction to Forensic Anthropology: A Textbook*. Boston, MA: Allyn & Bacon.

Claes, P., G. Willems, P. Suetens, D. Vandermeulen, and S. De Greef. 2006.

Craniofacial reconstruction using a combined statistical model of face shape and soft tissue depths: methodology and validation. *Forensic Science International* 159 Suppl 1:S147-58.

Clement, J. G., and M. K. Marks. 2005. "Introduction to Facial Reconstruction," in *Computer-Graphic Facial Reconstruction*. Edited by J. G. Clement and M. K. Marks, pp. 3-13. New York: Elsevier Academic Press.

Clement, J. G., and D. R. Ranson. 1998. "Craniofacial Identification," in *Craniofacial Identification in Forensic Medicine*. Edited by J. G. Clement and D. R. Ranson, pp. 3-8. London: Arnold.

Craig, E. A. 2003. Commentary on: Stephan CN. Facial approximation: globe projection guideline falsified by exophtalmometry literature. J Forensic Sci 2002;47(4):730-5. *J Forensic Sci* 48:469;author.

DeGreef, S., and G. Willems. 2005. Three-Dimensional Cranio-Facial Reconstruction in Forensic Identification: Latest Progress and New Tendencies in the 21st century. *Journal of Forensic Sciences* 50:12-17.

DeGreef, S., P. Claes, D. Vandermeulen, W. Mollemans, P. Suetens, and G. Willems. 2006. Large-scale in-vivo Caucasian facial soft tissue thickness database for craniofacial reconstruction. *Forensic Science International* 159:S126-S146.

Domaracki, M., and C. N. Stephan. 2006. Facial Soft Tissue Thicknesses in Australian Adult Cadavers. *Journal of Forensic Sciences* 51(1):5-10.

Doty, N. D. 1998. The Influence of nationality on the Accuracy of Face and Voice Recognition. *The American Journal of Psychology* 111:191-214.

El-Mehallawi, I. H., and E. M. Soliman. 2001. Ultrasonic assessment of facial soft tissue thicknesses in adult Egyptians. *Forensic Sci Int* 117:99-107.

1999. *Eyewitness Evidence: A Guide for Law Enforcement*. U. S. Department of Justice, Office of Justice Programs.

Evenhouse, R. M., M. Rasmussen, and L. Sadler. 1992. Computer-Aided Forensic Facial Reconstruction. *Journal of Biocommunication* 19:22-28.

Evison, M. P. 1996. "Computerised 3D Facial Reconstruction," in *Assemblage*, 1996(1) edition.

Evison, M. P. 2000. "Modeling Age, Obesity, and Ethnicity in Computerised 3D Facial Reconstruction." *9th International Craniofacial Identification Conference., Washington, DC, 2000*. Forensic Science Communications.

Evison, M. P., O. M. Finegan, and T. C. Blythe. 1998. "Computerised 3D Facial Reconstruction: Research Update," in *Assemblage*, 1998(4) edition.

Evison, M. P., and M. A. Green. 1999. Presenting three-dimensional forensic facial simulations on the Internet using VRML. *Journal of Forensic Sciences* 44:1219-1223.

1999. *Eyewitness Evidence: A Guide for Law Enforcement*. U. S. Department of Justice, Office of Justice Programs.

2006. "Eyewitness Identification." *Training Key® #600*: International Association of Chiefs of Police, Inc.

Friend, C. E. 1992. "Showups, Lineups and Photo Identifications." *Training Key® #414*: International Association of Chiefs of Police, Inc.

Gatliff, B. P. 1984. Facial Sculpture on the Skull for Identification. *American Journal of Forensic Medicine and Pathology* 5:327-332.

Gatliff, B. P. 1986. "Forensic Sculpture Adapts to Museum Use." *7th Annual Conference of the Build of Natural Science Illustraton, Washington, D.C., 1986*, pp. 13-17. Selected papers from the 7th Annual Conference of the Guild of Natural Science Illustrators.

Gatliff, B. P., and C. C. Snow. 1979. From Skull to Visage. *Journal of Biocommunication* 6:27-30.

Gatliff, B. P., and K. T. Taylor. 2001. "Three-Dimensional facial Reconstruction on the Skull," in *Forensic Art and Illustration*, pp. 419-475. Boca Raton, FL: CRC Press.

George, R. M. 1987. The Lateral Craniographic Method of Facial Reconstruction. *Journal of Forensic Sciences* 32:1305-1303.

George, R. M. 1993. "Anatomical and Artistic Guidelines for Forensic Facial Reconstruction," in *Forensic Analysis of the Skull: Craniofacial Analysis, Reconstruction, and Identification*. Edited by M. Y. Iscan and R. P. Helmer, pp. 215-227. New York: Wiley-Liss.

Gerasimov, M. M. 1971. *The Face Finder*, English Translation Copyright 1971 by Hutchinson & Co (Publishers) Ltd edition. Philadelphia and New York: J. B. Lippincott Company.

Green, M. A., and M. P. Evison. 1999. Interpolating between computerized three-dimensional forensic facial simulations. *Journal of Forensic Sciences* 44:1224-1228.

Gruner, O. 1993. "Identification of Skulls. A Historical Review and Practical Applications," in *Forensic Analysis of the Skull: Craniofacial Analysis, Reconstruction and Identification*. Edited by M. Y. Iscan and R. P. Helmer, pp. 29-45. New York: Wiley-Liss.

Haglund, W. 1998. "Forensic 'art' in human identification.," in *Craniofacial Identification in Forensic Medicine*. Edited by J. G. R. Clement, David L., pp. 235-243. London: Arnold.

Haglund, W. D., and D. T. Reay. 1991. Use of Facial Approximation in Identification of Green River Serial Murder Victims. *American Journal of Forensic Medicine and Pathology* 12:132-142.

Hancock, P. J. B., V. Bruce, and A. M. Burton. 2000. Recognition of Unfamiliar Faces. *Trends in Cognitive Sciences* 4:330-337.

Helmer, R. P., S. Rohricht, D. Peterson, and F. Mohr. 1993. "Assessment of the Reliability of Facial Reconstruction," in *Forensic Analysis of the Skull: Craniofacial Analysis, Reconstruction, and Identification*. Edited by M. Y. Iscan and R. P. Helmer, pp. 229-246. New York: Wiley-Liss.

Hering, P. 2003. "Ultra-fast three-dimensional facial profile measurement through pulsed holography for forensic facial reconstruction (Abstract)." *1st International Conference on Reconstruction of Soft Facial Parts, Potsdam, Germany, 2003*.

Hill, H., and V. Bruce. 1996. Effects of lighting on the perception of facial surfaces. *J Exp Psychol Hum Percept Perform* 22:986-1004.

His, W. 1895. *Anatomische Forschungen über Johann Sebastian Bach's Gebeine und Antlitz nebst Bemerkungen über dessen Bilder. Abhandlungen der Mahematisch-Physischen Classe der Königl. Sächsischen Gesellschaft der Wissenschaften*. Leipzig: Bei S. Hirzel.

Iscan, M. Y. 1988. Rise of Forensic Anthropology. *Yearbook of Physical Anthropology* 31:203-230.

Iscan, M. Y. 1993. "Craniofacial Image Analysis and Reconstruction," in *Forensic Analysis of the Skull: Craniofacial Analysis, Reconstruction, and Identification*. Edited by M. Y. Iscan and R. P. Helmer, pp. 1-9. New York: Wiley-Liss.

Iscan, M. Y. 1998. Progress in forensic anthropology: the 20th century. *Forensic Science International* 98:1-8.

Iscan, M. Y. 2001. Global forensic anthropology in the 21st century. *Forensic Science International* 117:1-6.

Iscan, M. Y., and R. P. Helmer. Editors. 1993. *Forensic Analysis of the Skull. Craniofacial Analysis, Reconstruction, and Identification*. New York: Wiley & Sons.

Iscan, M. Y., and G. Quatrehomme. 1999. Medicolegal anthropology in France. *Forensic Science International* 100:17-35.

Jones, M. W. 2001. "Facial Reconstruction Using Volumetric Data." *Proceedings of the 6th International Vision Modelling and Visualisation Conference, Stuttgart, Germany, 2001*.

Joyce, C. , and E. Stover 1991. *Witnesses from the Grave*. Boston: Little, Brown, and Company.

Kahler, K., J. Haber, and H. P. Seidel. 2003. Reanimating the Dead: Reconstruction of Expressive Faces from Skull Data. *ACM Transactions on Graphics* 22:554-561.

Kim, K. D., A. Ruprecht, G. Wang, J. B. Lee, D. V. Dawson, and M. W. Vannier. 2005. Accuracy of facial soft tissue thickness measurements in personal computer-based multiplanar reconstructed computed tomographic images. *Forensic Science International* 155:28-34.

Kollman, J., and W. Büchly. 1898. Die Persistenz der Rassen und die Reconstruction der Physiognomic prähistorischer Schädel. *Archiv fur Anthropologie* 25:329-359.

Krogman, W. M. 1962. *The Human Skeleton in Forensic Medicine*. Springfield, IL: Charles C. Thomas.

Lambrecht, J. T., F. Brix, and H. Gremmel. 1993. "Three-Dimensional Skull Identification via Computed Tomographic Data and Video Visualization," in *Forensic Analysis of the Skull: Craniofacial Analysis, Reconstruction, and Identification*. Edited by M. Y. Iscan and R. P. Helmer, pp. 97-104. New York: Wiley-Liss.

Linney, A., and A. M. Coombes. 1998. "Computer Modelling of Facial Form," in *Craniofacial Identification in Forensic Medicine*. Edited by J. G. Clement and D. R. Ranson, pp. 187-199. London: Arnold.

Manhein, M. H. 1999. *The Bone Lady: Life as a Forensic Anthropologist*. Baton Rouge, LA: Louisiana State University Press.

Manhein, M. H., G. A. Listi, R. E. Barsley, R. Musselman, N. E. Barrow, and D. Ubelaker. 2000. In vivo facial tissue-depth measurements for children and adults. *Journal of Forensic Sciences* 45:48-60.

Maples, W. R. 1994. *Dead Men Do Tell Tales: The Strange and Fascinating Cases of a Forensic Anthropologist*. New York: Doubleday.

Marks, M. K., J. L. Bennett, and O. L. Wilson. 1997. Digital Video Image Capture in Establishing Positive Identification. *Journal of Forensic Sciences* 42:492-495.

Marks, M. K., D. R. Tufano, E. C. Uberbacher, R. E. Flanery, V. N. Olman, and Y. Xu. 2000. "Artificial Intelligence/Computational Method to Facial Reconstruction (Abstract)." 9th Biennial Scientific Meeting of the International Association for Craniofacial Identification Conference, Federal Bureau of Investigation, Washington, DC, 2000. Forensic Science Communications 2(4).

Megreya, A. M., and A. M. Burton. 2006. Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition* 34:865-876.

Miller, K. W. P., M. A. Taister, T. P. Kelliher, and P. H. Tu. 2003. Facial Reconstruction of Human Remains Using Deformation Modes. *Forensic Science International* 136:179.

Milner, C., C. M. Wilkinson, and R. Neave. 2000. "Adult Craniofacial Aging an Craniofacial Reconstruction." 9th Biennial Scientific Meeting of the International Association for Craniofacial Identification Conference, Federal Bureau of Investigation, Washington, DC, 2000. Forensic Science Communications 2(4).

Mittleman, R. E., R. Souviron, J. Valor, and L. Lugo. 1983. "Facial Reconstruction, Publicity, and Photography Leading to Identification of Deceased Persons," in *FBI Law Enforcement Bulletin*, vol. 52, pp. 21-23: Office of Congressional and Public Affairs.

Miyasaka, S., M. Yoshimo, K. Imaizumi, and S. Seta. 1995. The computer-aided facial reconstruction system. *Forensic Science International* 74:155-165.

Moss, J. P., A. D. Linney, S. R. Grindrod, and C. A. Mosse. 1989. A Laser Scanning System for the Measurement of Facial Surface Morphology. *Optics and Lasers in Engineering* 10:179-190.

Nelson, L. A., and S. D. Michael. 1998. The application of volume deformation to three-dimensional facial reconstruction: A comparison with previous techniques. *Forensic Science International* 94:167-181.

Phillips, V. M. 2000. "Skeletal Remains Identification by Facial Reconstruction." *Proceedings of the 9th Biennial Meeting of the International Association for Craniofacial Identification, Washington, DC, 2000*. Forensic Science Communications 3.

Phillips, V. M., and N. A. Smuts. 1996. Facial Reconstruction: Utilization of Computerized Tomography to Measure Facial Tissue Thickness in a Mixed Racial Population. *Forensic Science International* 83:51-59.

Pourtois, G., S. Schwartz, M. L. Seghier, F. Lazeyras, and P. Vuilleumier. 2005. Portraits or people? Distinct representations of face identity in the human visual cortex. *Journal of Cognitive Neuroscience* 17:1043-1057.

Prag, J. N., and R. A. H. Neave. 1997. *Making Faces: Using Forensic and Archeological Evidence*. London: British Museum Press.

Quatrehomme, G., and S. Cotin, et al. 1997. A fully three-dimensional method for facial reconstruction based on deformable models. *Journal of Forensic Sciences* 42:649-652.

Reichs, K. J., and E. A. Craig. 1998. "Facial Approximation: Procedures and Pitfalls," in

Forensic Osteology: Advances in the Identification of Human Remains, 2nd edition. Edited by K. J. Reichs, pp. 491-513. Springfield, IL: Charles C. Thomas.

Rhine, J. S. 1990. Coming to Terms with Facial Reproduction. *J Forensic Sci* 35:960-3.

Rhine, J. S. 1998. *Bone Voyage*. Albuquerque, NM: University of New Mexico Press.

Rhine, J. S., and H. R. Campbell. 1980. Thickness of Facial Tissues in American Blacks. *Journal of Forensic Sciences* 25:847-858.

Ritter, N. 2007. Missing Persons and Unidentified Remains: The Nation's Silent Mass Disaster. *National Institute of Justice Journal* 256:2-7.

Rocha, S. d. S., D. L. d. P. Ramos, and M. d. G. P. Cavalcanti. 2003. Applicability of 3D-CT Facial Reconstruction for Forensic Individual Identification. *Pesquia Odontologica Brassileira (Brazilian Oral Research)* 17:24-28.

Sahni, D., I. Jit, et al. 2002. "Preliminary Study on Facial Soft Tissue Thickness by Magnetic Resonance Imaging in Northwest Indians." *Forensic Science Communications* 4(1).

Schwartz, J. H. 1993. *What the Bones Tell Us*. New York: Henry Holt and Company.

Shahrom, A. W., P. Vanezis, R. C. Chapman, A. Gonzales, C. Blenkinsop, and M. L. Rossi. 1996. Techniques in Facial Identification: Computer-aided Facial Reconstruction Using Laser Scanner and Video Superimposition. *International Journal of Legal Medicine* 108:194-200.

Simpson, E., and M. Henneberg. 2002. Variation in soft-tissue thicknesses on the human face and their relation to craniometric dimensions. *Am J Phys Anthropol* 118:121-33.

Smith, S. L., and G. S. Throckmorton. 2004. A New Technique for Three-Dimensional Ultrasound Scanning of Facial Tissues. *Journal of Forensic Sciences* 49:451-457.

Snow, C. C., B. P. Gatliff, and K. R. McWilliams. 1970. Reconstruction of Facial Features from the Skull: An Evaluation of its Usefulness in Forensic Anthropology. *American Journal of Physical Anthropology.* 33:221-227.

Stephan, C. N. 2002a. Do resemblance ratings measure the accuracy of facial approximations? *Journal of Forensic Sciences* 47:239-43.

Stephan, C. N. 2002b. Position of superciliare in relation to the lateral iris: testing a suggested facial approximation guideline. *Forensic Science International* 130:29.

Stephan, C. N. 2002c. Facial approximation: Globe projection guideline falsified by exophthalmometry literature. *Journal of Forensic Sciences* 47:730-735.

Stephan, C. N. 2003a. Anthropological facial 'reconstruction--recognizing the fallacies, 'unembracing' the errors, and realizing method limits. *Sci Justice* 43:193-200.

Stephan, C. N. 2003b. Facial Approximation: An Evaluation of Mouth-Width Determination. *American Journal of Physical Anthropology* 121:48.

Stephan, C. N. 2006. Beyond the sphere of the English facial approximation literature: Ramifications of German papers on western method concepts. *Journal of Forensic Sciences* 51:736-739.

Stephan C. N. and R.S. Arthur. 2006. Assessing facial approximation accuracy: How do resemblance ratings of disparate faces compare to recognition tests? *Forensic Science International* 159: S159-S63

Stephan, C. N., and M. Henneberg. 2001. Building faces from dry skulls: are they recognized above chance rates? *Journal of Forensic Sciences* 46:432-40.

Stephan, C. N., and M. Henneberg. 2006. Recognition by forensic facial approximation: case specific examples and empirical tests. *Forensic Sci Int* 156:182-91.

Stephan, C. N., M. Henneberg, and W. Sampson. 2003. Predicting nose projection and pronasale position in facial approximation: a test of published methods and proposal of new guidelines. *Am J Phys Anthropol* 122:240-50.

Stoney, M. B., and T. D. Koelmeyer. 1999. Facial reconstruction: a case report and review of development of techniques. *Medicine, Science, and the Law* 39:49-60.

Suzuki, K. 1948. On the Thickness of the Soft Parts of the Japanese Face. Journal of the Anthropological Society of Nippon *60:*7-11.

Taylor, K. T. 2001. *Forensic Art and Illustration*. Boca Raton, FL: CRC Press.

Taylor, J. A., and K. A. Brown. 1993. "Superimposition Techniques," in *Craniofacial Identification in Forensic Medicine*. Edited by J. G. Clement and D. R. Ranson, pp. 151-164. London: Arnold.

Taylor, R. G., and C. Angel. 1998. "Facial Reconstruction and Approximation," in *Craniofacial Identification in Forensic Medicine*. Edited by J. G. Clement and D. R. Ranson, pp. 177-185. London: Arnold.

Tu, P. H., R. I. Hartley, W. E. Lorensen, M. Allyassin, R. Gupta, and L. Heier. 2000. Face Reconstruction Using Flesh Deformation Models (Abstract). *Forensic Science Communications* 2.

Tu, P. H., T. P. Kelliher, K. W. P. Miller, and M. A. Taister. 2003. Towards a Statistical Basis for Facial Deformation Modes in Reconstruction. *Forensic Sci Int* 136:168-169.

Tu, P. H., R. I. Hartley, W. E. Lorensen, A. Alyassin, R. Gupta, and L. Heier. 2005.

"Face Reconstructions Using Flesh Deformation Modes," in *Computer-Graphic Facial Reconstruction*. Edited by M. K. Marks and J. G. Clement, pp. 143-160. TBA: (In Press).

Turner, W. D., R. E. B. Brown, T. P. Kelliher, P. H. Tu, M. A. Taister, and K. W. P Miller. 2005. A novel method of automated skull registration for forensic facial approximation. *Forensic Science International* 154:149-158.

Turner, W., P. Tu, T. Kelliher, and R. Brown. 2006. Computer-Aided Forensics: Facial Reconstruction. *Stud Health Technol Inform* 119:550-5.

Tyrrell, A. J., M. A. Green, M. P. Evison, and A. T. Chamberlain. 1997. Forensic Three-Dimensional Facial Reconstruction: Historical Review and Contemporary Developments. *Journal of Forensic Sciences* 42:653-61.

Ubelaker, D. H. 2000. "A History of Smithsonian-FBI Collaboration in Forensic Anthropology, Especially in Regard to Facial Imagery." *9th Biennial Meeting of the International Association for Craniofacial Identification, October 2000, FBI, Washington, D.C., 2000* 2(4).

Ubelaker, D. H., and G. O'Donnell. 1992. Computer-Assisted Facial Reproduction. *Journal of Forensic Sciences* 37(1):155-162.

Ubelaker, D. H., and H. Scammell. 1992. *Bones, a Forensic Detective's Casebook*. New York: M. Evans and Company.

Valentin, D., H. Abdi, and B. Edelman. 1997. What represents a face? A computational approach for the integration of physiological and psychological data. *Perception* 26:1271-1288.

Valentine, T., and V. Bruce. 1986. The effect of race, inversion and encoding activity upon face recognition. *Acta Psychological* 61:259-273.

Vandermeulen, D., P. Claes, D. Loeckx, S. De Greef, G. Willems, and P. Suetens. 2006. Computerized craniofacial reconstruction using CT-derived implicit surface representations. *Forensic Science International* 159:S164-S174.

Vanezis, M., and P. Vanezis. 2000. Cranio-Facial Reconstruction in Forensic Identification - Historical Developments and a Review of Current Practice. *Medicine, Science, and the Law* 40:197-205.

Vanezis, P., R. W. Blowes, A. D. Linney, A. C. Tan, R. Richards, and R. Neave. 1989. Application of 3-D Computer Graphics for Facial Reconstruction and Comparison with Sculpting Techniques. *Forensic Science International* 42:69-84.

Vanezis, P., M. Vanezis, G. McCombe, and T. Niblett. 2000. Facial reconstruction using 3-D computer graphics. *Forensic Science International* 108:81-95.

Wallach, H., and D. N. O'Connell. 1953. The Kinetic Depth Effect. *Journal of*

Experimental Psychology 45(4):205-217.

Wilkinson, C. M. 2002. In Vivo Facial Tissue Depth Measurements for White British Children. *Journal of Forensic Sciences* 47:459-465.

Wilkinson, C. M. 2004a. Facial approximation: Comments on Stephan (2003). *American Journal of Physical Anthropology* 125:329-330.

Wilkinson, C. M. 2004b. *Forensic Facial Reconstruction*. Cambridge, UK: Cambridge University Press.

Wilkinson, C. M., R. A. H. Neave, D. Smith, and B. A. Hons. 2002. "How Important to Facial Reconstruction are the Correct Ethnic Group Tissue Depths?" *A Head Start on the New Millennium: the Latest in Craniofacial Identification. Proceedings of the 10th Biennial Scientific Meeting of the International Association for Craniofacial Identification., Bari, Italy, 2002*, pp. 111-121.

Wilkinson, C., C. Rynn, H. Peters, M. A. Taister, C.H. Kau, and S. Richmond. 2006 A Blind Accuracy Assessment of Computer-Modeled Forensic Facial Reconstruction Using Computed Tomography Data from Live Subjects. Forensic Science, Medicine, and Pathology 2:3:179-187.

Wilkinson, C. M., and D. K. Whittaker. 2002. "Juvenile Forensic Facial Reconstruction - A Detailed Accuracy Study." *A Head Start on the New Millennium: The Latest in Craniofacial Identification. Proceedings of the 10th Biennial Scientific Meeting of the International Association for Craniofacial Identification, Bari, Italy, 2002*, pp. 98-110.

Williamson, M. A., S. P. Nawrocki, and T. A. Rathburn. 2002. Variation in Midfacial Tissue Thickness of African-American children. *Journal of Forensic Sciences* 47:25-31.

Wingate, A. 1992. *Scene of the Crime*. Cincinnati, OH: Writer's Digest Books.

Zonderman, J. 1999. *Beyond the Crime Lab. The New Science of Investigation*. New York: John Wiley & Sons, Inc.

Appendix A
Phase 2 Validation Testing materials

PARTICIPANT INSTRUCTIONS

Instructions to Validation Participants

Some of the photographs used in this research project may be from actual forensic cases. If any of the subjects in the photographs presented in this validation test are known to you, then you may not participate in this validation. In this event, please advise the administrator of this test and return all testing materials.

1. Please do not open any envelopes until requested to do so by the administrator.
2. Please answer all questions; an incomplete answer sheet can not be included in this project.
3. Please do not discuss your answers with anyone else taking the test.
4. You may not keep or remove any part of the test or test materials.

When instructed to do so by the administrator, open the **"Validation ONE Face-Pool"** envelope and remove the contents. DO NOT OPEN ANY ENVELOPES until requested to do so by the administrator.

C:\Documents and Settings\dkmoyers\DIANA\ReFace Test Protocol\Final Test Instructions.wpd

Appendix A-1:Handout of general instructions for participants in *Phase 2* photographic validation.

Validation ONE

Face-Pool

Work on only one face-pool validation test at a time. Do not open another validation test until the previous one has been completed and returned to the "Validation ONE Face-Pool" envelope. Once a face- pool validation test has been completed, you may not return to it.

1. Please remove the sealed envelopes and Answer Sheet inside this envelope. Verify that you have 10 sealed envelopes and an Answer Sheet.

2. The sealed envelopes have been randomly sorted. Please remove the top envelope and open it.

3. Please make your selection and mark your answer on the appropriate section of your answer sheet.

   NOTE: Your Face-Pool validation tests are not in the same order as your answer sheet. Please be sure to write your answer in the correct SUBJECT and Group section on the answer sheet.

4. Place the completed face-pool validation test back in its envelope and return it to the **"Validation ONE Face- Pool"** envelope.

5. Open the next sealed envelope, and repeat steps 3 and 4.

6. Continue opening the next sealed envelope, repeating steps 3 and 4, until there are no more envelopes.

7. Place all completed Validation ONE materials, including your Answer Sheet, inside the **"Validation ONE Face- Pool"** envelope and return it to the administrator.

This concludes Validation ONE. The administrator will give you the second part of the Validation. Please follow the written instructions for Validation TWO.

C:\Documents and Settings\dkmoyers\DIANA\ReFace Test Protocol\Final Test Instructions.wpd

Appendix A-2: Participant instructions for Validation ONE: Face-Pool.

ANSWER SHEET: Validation ONE (Face-Pool) **ENVELOPE NUMBER:** _______
Page One of Two

Sex: Male Female (please circle one) Age: __________

Please circle the response that best indicates your level of knowledge in the area of forensic facial approximation/reconstruction.

Very Knowledgeable	Somewhat Knowledgeable	No Prior Knowledge

SUBJECT A
White Females (Group 1)

Please circle the number of the photograph that most resembles Subject A

1 2 3 4 5 6 7

SUBJECT B
White Females (Group 1)

Please circle the number of the photograph that most resembles Subject B

1 2 3 4 5 6 7

SUBJECT C
White Females (Group 2)

Please circle the number of the photograph that most resembles Subject C

1 2 3 4 5 6 7

SUBJECT D
White Females (Group 2)

Please circle the number of the photograph that most resembles Subject D

1 2 3 4 5 6 7

SUBJECT E
White Females (Group 1)

Please circle the number of the photograph that most resembles Subject E

1 2 3 4 5 6 7

Appendix A-3: Face-Pool participant answer sheet (page 1 of 2).

ANSWER SHEET: Validation ONE (Face-Pool) **ENVELOPE NUMBER:** _______
Page Two of Two

==

SUBJECT F
White Females (Group 1)

Please circle the number of the photograph that most resembles Subject F

1 2 3 4 5 6 7

==

SUBJECT G
White Males (Group 1)

Please circle the number of the photograph that most resembles Subject G

1 2 3 4 5 6 7

==

SUBJECT H
White Males (Group 1)

Please circle the number of the photograph that most resembles Subject H

1 2 3 4 5 6 7

==

SUBJECT I
White Males (Group 1)

Please circle the number of the photograph that most resembles Subject I

1 2 3 4 5 6 7

==

SUBJECT J
White Males (Group 1)

Please circle the number of the photograph that most resembles Subject J

1 2 3 4 5 6 7

==

Appendix A-3 (continued): Face-Pool participant answer sheet (page 2 of 2).

Validation TWO

Resemblance Rating

Work on only one rating test at a time. Do not open another rating test until the previous rating test has been completed and returned to the "Validation TWO Resemblance Rating" envelope. Once a rating test has been completed, you may not return to it.

1. Please remove the sealed envelopes and Answer Sheet inside this envelope. Verify that you have ten sealed envelopes and an Answer Sheet.
2. The sealed envelopes have been randomly sorted. Please remove the top envelope and open it.
3. Please make your selection and mark your answer on the appropriate section of your answer sheet.

   NOTE: Your rating tests are not in the same order as your answer sheet. Please be sure to note your answer in the correct SUBJECT and Group on the answer sheet.
4. Place each completed rating test back in its envelope and return it to the "**Validation TWO Resemblance Rating**" envelope.
5. Open the next sealed envelope, and repeat steps 3 and 4.
6. Continue opening the next sealed envelope, repeating steps 3 and 4, until there are no more envelopes.
7. Place all completed Validation TWO materials, including your Answer Sheet, inside the "**Validation TWO Resemblance Rating**"envelope and return it to the administrator.

This concludes Validation TWO. Thank you for your participation.

C:\Documents and Settings\dkmoyers\DIANA\ReFace Test Protocol\Final Test Instructions.wpd

Appendix A-4: Participant instructions for Validation TWO: Resemblance Rating.

ANSWER SHEET: Validation TWO (Resemblance Rating) **ENVELOPE NUMBER:** _______

Page One of Two

==

Sex: Male Female (please circle your response) Age:_____________

==

Please circle the response that best indicates your level of knowledge in the area of forensic facial approximation/reconstruction.

Very Knowledgeable	Somewhat Knowledgeable	No Prior Knowledge

For each subject, please indicate how much the facial approximation on the left resembles the photograph on the right. (Please circle your response)

SUBJECT A
White Females (Group 1)

No Resemblance	Slight Resemblance	Approximate Resemblance	Close Resemblance	Strong Resemblance
1	2	3	4	5

SUBJECT B
White Females (Group 1)

No Resemblance	Slight Resemblance	Approximate Resemblance	Close Resemblance	Strong Resemblance
1	2	3	4	5

SUBJECT C
White Females (Group 2)

No Resemblance	Slight Resemblance	Approximate Resemblance	Close Resemblance	Strong Resemblance
1	2	3	4	5

SUBJECT D
White Females (Group 2)

No Resemblance	Slight Resemblance	Approximate Resemblance	Close Resemblance	Strong Resemblance
1	2	3	4	5

SUBJECT E
White Females (Group 1)

No Resemblance	Slight Resemblance	Approximate Resemblance	Close Resemblance	Strong Resemblance
1	2	3	4	5

Appendix A-5: Resemblance Rating participant answer sheet (page 1 of 2).

ANSWER SHEET: Validation TWO (Resemblance Rating) **ENVELOPE NUMBER:** _______
Page Two of Two

For each subject, please indicate how much the facial approximation on the left resembles the photograph on the right. (Please circle your response)

SUBJECT F
White Females (Group 1)

No Resemblance	Slight Resemblance	Approximate Resemblance	Close Resemblance	Strong Resemblance
1	2	3	4	5

SUBJECT G
White Males (Group 1)

No Resemblance	Slight Resemblance	Approximate Resemblance	Close Resemblance	Strong Resemblance
1	2	3	4	5

SUBJECT H
White Males (Group 1)

No Resemblance	Slight Resemblance	Approximate Resemblance	Close Resemblance	Strong Resemblance
1	2	3	4	5

SUBJECT I
White Males(Group 1)

No Resemblance	Slight Resemblance	Approximate Resemblance	Close Resemblance	Strong Resemblance
1	2	3	4	5

SUBJECT J
White Males (Group 1)

No Resemblance	Slight Resemblance	Approximate Resemblance	Close Resemblance	Strong Resemblance
1	2	3	4	5

Appendix A-5 (continued): Resemblance Rating participant answer sheet (page 2 of 2).

Appendix B

Demographics of Test Subjects and ReFace Database

UTK Validation Subjects Demographics: All Populations					
Cumulative	SEX	Scan ID	Age	Height (in.)	Weight (lb)
1	Female	11-04D	54	65	120
2	Female	12-04D	60	66	125
3	Female	18-04D	44	69	350
4	Female	26-04D	69	61	0
5	Female	17-03D	58	62	101
6	Female	18-03D	47	61	103
7	Female	43-03D	73	67	150
8	Female	08-02D	79	0	155
9	Female	35-02D	55	69	139
10	Female	37-02D	52	64	183
11	Female	03-01D	62	0	104
12	Female	15-01D	61	61	200
13	Female	28-01D	61	65	200
14	Female	99-03F	15	0	0
15	Female	96-10F	28	66	170
16	Female	89-13F	16	0	0
17	Female	85-30F	18	65	0
18	Female	81-11F	25	66	0
19	Female	79-10F	58	65	0
20	Male	15-05D	53	68	172
21	Male	42-05D	42	74	175
22	Male	46-05D	60	69	165
23	Male	06-04D	16	0	50
24	Male	24-04D	61	0	210
25	Male	29-04D	34	73	220
26	Male	36-04D	83	64	150
27	Male	39-04D	60	67	100
28	Male	43-04D	72	72	150
29	Male	31-03D	88	67	125
30	Male	49-03D	86	71	150
31	Male	52-03D	55	74	201
32	Male	60-03D	79	70	185
33	Male	01-02D	96	66	175
34	Male	10-02D	61	0	192
35	Male	11-02D	76	0	170
36	Male	20-02D	65	69	150
37	Male	09-01D	44	73	225
38	Male	21-01D	69	70	163
39	Male	29-01D	84	65	134
40	Male	31-01D	77	0	175
41	Male	42-01D	76	70	278
42	Male	48-01D	59	68	0
43	Male	05-98D	67	68	0
44	Male	18-93D	78	0	0
45	Male	14-92D	57	0	0
46	Male	02-89D	36	72	220
47	Male	07-87D	58	67	138
48	Male	86-18F	26	70	145
49	Male	81-09F	25	71	140
50	Black Female	85-16F	29	60	0
51	Black Male	21-92D	25	75	175
52	Black Male	96-30F	35	67	0
53	Black Male	86-30F	73	71	0
UTK Subjects Statistical Summary (N = 53)					
Cumulative	SEX	Scan ID	Age	Height (in.)	Weight (lb)
		Mean	55.47	67.69	165.83
		Median	59	67.5	164
		Mode	61	65	150
	Range	Minimum	15	60	50
	Range	Maximun	96	75	350

Appendix B-6: University of Tennessee Test Subject Demographics (N = 53)

UTK Subjects Statistical Data				
Age Distribution: American Females: European Descent				
Cumulative	Scan ID	Age	Height (in.)	Weight (lb)
1	11-04D	54	65	120
2	12-04D	60	66	125
3	18-04D	44	69	350
4	26-04D	69	61	0
5	17-03D	58	62	101
6	18-03D	47	61	103
7	43-03D	73	67	150
8	08-02D	79	0	155
9	35-02D	55	69	139
10	37-02D	52	64	183
11	03-01D	62	0	104
12	15-01D	61	61	200
13	28-01D	61	65	200
14	99-03F	15	0	0
15	96-10F	28	66	170
16	89-13F	16	0	0
17	85-30F	18	65	0
18	81-11F	25	66	0
19	79-10F	58	65	0
	Mean	49	65	162
	Median	55	65	150
	Mode	58	65	200
Range	Minimum	15	61	101
Range	Maximun	79	69	350

Appendix B-7:University of Tennessee Validation Subjects Demographics: American Females: European Descent (N = 19).

UTK Validation Subjects Statistical Data				
Age Distribution: American Males: European Descent				
Cumulative	Scan ID	Age	Height (in.)	Weight (lb)
1	15-05D	53	68	172
2	42-05D	42	74	175
3	46-05D	60	69	165
4	06-04D	16	0	50
5	24-04D	61	0	210
6	29-04D	34	73	220
7	36-04D	83	64	150
8	39-04D	60	67	100
9	43-04D	72	72	150
10	31-03D	88	67	125
11	49-03D	86	71	150
12	52-03D	55	74	201
13	60-03D	79	70	185
14	01-02D	96	66	175
15	10-02D	61	0	192
16	11-02D	76	0	170
17	20-02D	65	69	150
18	09-01D	44	73	225
19	21-01D	69	70	163
20	29-01D	84	65	134
21	31-01D	77	0	175
22	42-01D	76	70	278
23	48-01D	59	68	0
24	05-98D	67	68	0
25	18-93D	78	0	0
26	14-92D	57	0	0
27	02-89D	36	72	220
28	07-87D	58	67	138
29	86-18F	26	70	145
30	81-09F	25	71	140
	Mean	61	69	168
	Median	61	70	167.5
	Mode	60	70	150
Range	Minimum	16	64	50
Range	Maximun	96	74	278

Appendix B-8: University of Tennessee Validation Subjects Demographics: American Males: European Descent.

ReFace Database Statistical Data				
Age Distribution: American Males: European Descent				
Cumulative	Scan ID	Age	Height (in.)	Weight (lb)
1	CAU 01	32	72	240
2	CAU 02	51	72	180
3	CAU 04	31	71	215
4	CAU 05	45	56	165
5	CAU 12	46	71	186
6	CAU 14	42	71	202
7	CAU 30	27	69	165
8	CAU 37	44	70	270
9	CAU 38	27	69	165
10	CAU 39	30	66	250
11	CAU 41	54	0	0
12	CAU 42	44	0	0
13	CAU 44	21	0	0
14	CAU 46	36	0	0
15	CAU 47	24	70	140
16	CAU 48	50	66	184
17	CAU 51	19	69	180
18	CAU 55	21	69	145
19	CAU 58	37	76	250
20	CAU 59	21	73	203
21	CAU 60	20	73	166
22	CAU 61	18	70	150
23	CAU 62	20	69	165
24	CAU 63	23	73	180
25	CAU 64	22	72	140
26	CAU 66	18	72	180
27	CAU 67	18	72	145
28	CAU 68	18	71	210
29	CAU 69	18	76	277
30	CAU 70	20	72	205
31	CAU 71	18	66	130
32	CAU 72	44	67	148
33	CAU 73	19	70	165
34	CAU 74	36	72	180
35	CAU 75	24	78	180
36	CAU 76	19	74	168
37	CAU 77	19	75	172
38	CAU 78	19	72	152
39	CAU 79	24	67	155
40	CAU 80	41	75	192
41	CAU 81	28	72	157
42	CAU 86	41	67	145
43	CAU 87	28	69	165
44	CAU 89	47	69	180
45	CAU 91	42	67	225
46	CAU 92	33	64	207
47	CAU 98	45	66	156
48	CAU 100	40	73	190
49	CAU 101	42	64	190
	Mean	30.73469388	70.15555556	182.5555556
	Median	28	71	180
	Mode	18	72	180
Range	Minimum	18	56	130
Range	Maximun	54	78	277

Appendix B-9: ReFace Statistics American Males (European Descent), N = 49).

ReFace Database Statistical Data				
Age Distribution: American Females: European Descent				
Cumulative	Scan ID	Age	Height (in.)	Weight (lb)
1	CAU 03	48	67	165
2	CAU 06	38	64	121
3	CAU 07	50	65	224
4	CAU 08	20	64	155
5	CAU 09	44	60	102
6	CAU 10	26	65	140
7	CAU 11	52	68	140
8	CAU 13	40	60	160
9	CAU 15	50	64	190
10	CAU 16	44	64	106
11	CAU 17	39	67	142
12	CAU 18	31	69	165
13	CAU 19	40	64	129
14	CAU 20	24	66	155
15	CAU 21	58	58	100
16	CAU 22	29	60	96
17	CAU 23	24	63	138
18	CAU 24	40	64	140
19	CAU 25	44	66	160
20	CAU 26	54	62	120
21	CAU 27	33	62	175
22	CAU 28	30	64	152
23	CAU 29	39	64	120
24	CAU 31	37	66	149
25	CAU 32	42	66	165
26	CAU 33	24	65	128
27	CAU 34	44	63	105
28	CAU 35	42	67	122
29	CAU 40	22	0	0
30	CAU 43	34	0	0
31	CAU 45	26	65	125
32	CAU 49	22	64	180
33	CAU 50	18	70	198
34	CAU 52	19	65	125
35	CAU 53	20	71	134
36	CAU 54	54	68	130
37	CAU 56	20	61	103
38	CAU 57	23	63	115
39	CAU 82	24	71	180
40	CAU 83	49	64	146
41	CAU 84	54	68	116
42	CAU 85	22	66	125
43	CAU 88	46	60	129
44	CAU 90	42	64	180
45	CAU 93	32	60	132
46	CAU 94	22	59	110
47	CAU 95	21	65	160
48	CAU 96	45	63	140
49	CAU 97	23	62	145
50	CAU 99	47	66	127
	Mean	35.42	64.4166667	140.9166667
	Median	37.5	64	139
	Mode	44	64	140
Range	Minimum	18	58	96
Range	Maximun	58	71	224

Appendix B-10: ReFace Statistics American Females (European Descent), N = 50).

Appendix C

Phase 1: Beta Version Software Validations

Appendix C
Phase 1: Beta Version Software Validations

The software evaluation of the facial reconstruction software began with Beta Version 1.0 and ended with Beta Version 5.2. Problems identified and corrected include (1) the import of dry bone, (2) protocol for the articulation of edentulous skulls for CT import, and (3) a method for the correction of extraneous bone or materials from packaging protocol. Facial approximations generated through Beta Version 5.2 were completely automated and produced an average face, although weight could be increased or decreased after the facial approximation was completed.

The prototype software operates in three distinct stages: (1) import CT data from an unknown skull into the program, (2) open workspace with a model skull created from import, and (3) build a facial approximation. The prototype testing results are organized in these categories.

ReFace Beta Version 1.0

Validation of the original prototype was scheduled to begin on May 16, 2005. However, the prototype consistently aborted before facial approximations were completed. No validation tests were possible until this error was corrected.

ReFace Validation I (Beta Version 2.0)

CT scans of 30 University of Tennessee subject skulls were used to test this version. A total of 92 tests were performed.

<u>Import Skulls</u>: (80 tests). There were 80 attempts (54 completed the import process; 26 failed to complete the import process). The Average import time was 20 seconds.

<u>Workspace</u>: (54 tests). There were 54 loads from previous imports. The only complete skull model was from the GE sample skull derived from a living person. There were no complete skull models produced from the University of Tennessee at Knoxville (UTK) sample; all workspace loads from UTK samples produced partial skulls (30 model skulls were ½ or less complete, usually only top of the skull); 11 skulls more than 50% but less than 75% (most were missing the maxilla and mandible, some were missing the entire face area); 12 skulls more than 75% but less than complete (maxilla and eye orbit areas were consistently missing, but present in “dots” when moving skull).

<u>Facial Approximations</u>: (14 tests). A total of 14 tests was conducted (four completed the process and ten failed). Of the four completed approximations, only the GE skull produced normal faces. The UTK approximations resulted in a “blob” woman and a “chin” man, with bone sticking outside the skin surface of the approximation. Approximation time was between 12 and 15 hours each.

Summary and Conclusions:
ReFace Beta Version 2.0 did not perform within expected parameters. CT data from skulls did not import properly and facial approximations produced from those skulls were incorrect. The only correct imports and approximations were produced from a "sample" skull from GE that was artificially created by removing the "skin" layer from the CT scan of a living person. This program was never tested with actual skeletal material; skull models created from living persons are not representative of skeletal material. It is possible the prototype would function as it is with fully articulated decomposing remains, but Beta Version 2.0 did not function with CT scans of actual skeletal remains.

ReFace Validation II (Beta Version 3.0)

Results from Validation I testing were relayed to the program developers and corrections were incorporated into Beta Version 3. Preliminary problem analysis found that the UTK scans were done at 5.0 mm, rather than 2.5 mm as requested. This was determined not to be the cause of the program malfunctions, but could produce results that were not as sufficiently detailed as with 2.5 mm scans. As the program was never tested with actual skeletal material, it was determined that the specifications for wet bone (from living persons) would not work with dry bone (skeletal material). A program correction was made and a new user interface added to indicate dry bone prior to import of the CT data. Validation Testing began on June 8, 2005 and ended on August 8, 2005. A total of 47 tests was performed.

Import Skulls: (37 tests). Imports of all subject skulls were conducted.
37 attempts (36 completed the import process; one failed to complete the import process). Import time ranged from a minimum of 9 seconds to a maximum of 35 seconds (mean is 11 seconds and mode is 11.5 seconds).

Workspace: (36 loads from previous imports.) Skull models in the workspace are complete skulls. There were three problems found in the workspace models which resulted in the elimination of 20 UTK skulls from the initial sample of 30.

1. *Wooden Dowels* used to attach and align the mandible with the cranium are faintly visible in dots in eight of the samples. Dowel slippage during scanning resulted in the mandible being out of alignment in three of the eight samples. One sample was still producing only half of the skull in the model. These eight samples could not be used for validation unless rescanned.

2. *Dental Noise* is present is nine samples. Five had severe dental noise with resulting spikes on the skull model; two had dental noise on only one side of the face; two had only slight dental noise and were possibly usable. Reconstructions were be done on the two samples with dental noise on one side only for testing and comparison of the differences in skin overlay by ReFace. Until a method is available to remove the dental noise, these samples could not be used for validation.

3. *CT Scan Errors* were found in three of the sample scans which resulted in incomplete skull models (chins cut off in two scans and edge of teeth cut off in one scan). These samples could not be used for validation unless rescanned.

Facial Reconstructions: Ten tests were conducted (eight completed the process and two failed). Approximation time was less than 3 hours each.

Case 35-02D: Power failure, error messages, and slight dental noise**.** (Four tests). Two failures and two completed approximations were from the same sample (35-02D). After import, workspace revealed slight dental noise. A reconstruction was attempted and failed due to a power outage during mapping of 48 out of 49 White database females. After the power was restored, an attempt to open the reconstruction resulted in an error message and system shutdown. A second attempt was made by opening the skull model and "Start Reconstruction" resulted in the reconstruction picking up from mapping 48 and finishing the reconstruction. Total time for the combined reconstructions: 3 hours 17 minutes. A second approximation was done to determine if the files had been damaged by the power failure. Results of both completed approximations were identical.

Case 11-04D: Visible Dowel and Mandible Alignment (two tests)
First import workspace had missing bone and a faint image of the wooden dowel. (The complete face was missing from original import.) A second import was done with a new name and case number. All comments written in the second import overwrote and appeared in all previous tests using this CT data. Reconstruction completed in 3 hours 17 minutes. (The resulting image had a skin "pouch" resulting from the dowel.) No further reconstructions were attempted if a dowel was visible in the workspace.

Case 21-92D: Black Male Approximations (two tests)
Only two Black males and nine black females were in the ReFace database as of July 12, 2005. Two approximations were done on the single UTK Black male sample skull for testing purposes only. The first approximation pulled from the correct population and sex database (two Black males). As expected, the approximation was very poor, with facial features poorly defined and blurred. A second test approximation was done without indicating whether the sample was male or female. ReFace correctly pulled all members of the Black population in the database (two males and nine females). Again, as expected, the approximation was very poor.

New Areas of Concern Identified with ReFace Beta 3.0

Edentulous Skulls: Wooden dowels were used to mimic the correct alignment for skulls without teeth. Previous imports were primarily the top half of the skull only. Now that the full skull was being produced, the wooden dowel used to align the mandible with the rest of the skull appeared on the skull model as a very faint image. They did not appear in the CT scans. When a reconstruction was attempted, the program thought this image was bone and put an extended "pouch" of skin on the cheek area. Current scans with the dowel exposed on one or both sides could not be used for validation testing until a method of removing the dowel image was found.

Additionally, some of the edentulous skulls (with and without a visible dowel in the workspace) were not correctly aligned due to dowel slippage and skull movement during CT scanning. The shorter the dowel, the greater the risk for slippage and thus mandibles that were not correctly aligned. Any current scans that were out of alignment could not be used for validation. However, these skulls could also be scanned again with the next sample. A protocol to better stabilize the skull during transport and scanning was needed.

CT imports with the wooden dowel cut to exactly the right size are visible only on the inside of the skull model and did not interfere with reconstructions. Future scans were corrected by cutting the proper length dowel. However, this did increase the preparation time for each skull prior to scanning and the possibility of slippage of the dowel, which resulted in improper alignment of the mandible. New materials may be found for dowels (such as plastic) that would not be interpreted by ReFace as bone. If a substitute that was "invisible" to the program was found, then longer dowels could be used to decrease the slippage problem.

Dental Noise: Most of the testing skulls had (and forensic skulls would be expected to have) dental fillings which produced dental noise on the CT scans, resulting in distortions of the skin overlay. ReFace did not remove this noise; some of the new Asian database skulls had severe dental noise which appeared on the CT skin image. This was a serious problem that would have to be addressed before ReFace could be fully operational.

Summary of Beta 3.0 Testing
Dental noise (9), dowel/alignments (8) and CT scan errors (3) reduced the original sample size of skulls from 30 to 10 (four White males, one Black male, and five White females). The CT errors and dowel problems could be corrected by rescanning the problem skulls.

ReFace Validation III (Beta Version 3.1)

This updated version contained no programming changes, but added 32 new database files and corrected 23 database files. Headholders (resulting from fabric on the CT head tray) were manually removed by the program developers from 22 existing files. Two files that were backward and upside down (CAU 95 and CAU 96) were corrected. These 24 files were activated in the database. There were 32 new database files (10 Black males and 23 Asians). Skin and bone models for all fixed and new files were visually verified.

Validation testing began on August 8, 2005 and ended on November 16, 2005. The entire validation sample of 30 UTK skulls was tested after installation of Beta Version 3.1. Validation testing was divided into four categories (regular sample, dental noise, dowel/alignment, and CT errors) and 71 tests were performed. Testing results indicated that ReFace appeared able to compensate if only one side of the face was affected (CT errors (one case returned to sample); Dowel/Alignment (one case returned to sample); Dental Noise (one case returned to sample and five potential cases identified for return to sample after retesting with Beta 4.0).

Results of Validation Test III (ReFace Beta Version 3.1)
Previous testing of ReFace Beta 3.0 (which produced complete skulls for the first time) identified problems with dental noise (nine cases), dowel/alignments (eight cases) and CT scan errors (three cases). Elimination of these cases reduced the original UTK sample of 30 skulls to ten skulls (four White males, one Black male, and five White females). All 30 UTK skulls and the GE skull were tested with ReFace Beta 3.1.

Test data from the GE wet skull were much higher in the import skull categories and test data from the single Black male subject skull (UTK Case 21-92D; with only 12 Black database males) were much lower in total time to complete a facial reconstruction as compared to the other test subjects. As these two extremes tended to distort the averages obtained, averages were provided with and without these two cases (outliers) in the overall and regular sample categories.

Summary of Outliers:

GE Wet Skull Approximation
Extreme differences in time for the import (69 seconds vs an average of 2.13 seconds) and building the bone model (20 seconds vs an average of 10.39 seconds) were noted. However, the reconstruction time (3 hours 22 minutes) was within the same range as the dry skulls.

Case 21-92D: Black Male Approximation
DICOM data copy was at the lowest end of the range (one second vs an average of 2.13 seconds). The bone model time was within the range of the other subjects (10 seconds vs an average of 10.39 seconds). The total time for import was also within the range of the other subjects (10 seconds vs an average of 12.52 seconds). The approximation time was the lowest (55 minutes vs an average of 3 hours 15 minutes) and believed to be due to the limited number of Black males in the database (12 as of February 8, 2006). ReFace correctly identified and pulled up all 12 members from the Black male population database. As expected, the approximation was poor, but showed definite improvement over the previous approximation with only two black males in the ReFace database.

New Information Discovered: The number of post-processing “Effects” equaled the number of “heads” in the population database.

Overall Testing Results (31 Subjects, 71 Tests)
A total of 71 tests was performed on 31 subject skulls after the update was applied. There were 35 import attempts of CT data from an unknown skull into the program with two failures. There were 33 attempted skull model openings in the workspace (three failed). There were 36 attempted facial approximations (two failed).

Import Skulls: (35 tests). New imports of all 30 UTK subject skulls were done under new case numbers. The GE wet skull was also tested. Results for the GE skull were much higher than the dry bone skulls and much lower for the single Black male subject (with only 12 CT database files). Averages without these two outliers were given first and averages including these two outliers followed. There were 35 import attempts (33 completed and two failed the import process). Total time to copy the DICOM data ranged from 0 seconds to 13 seconds with an average of 2.13 seconds. (Including the outliers, the range was from 0 seconds to 69 seconds with an average of 4.12 seconds.)

Creation of the bone model image (bone.vtk) ranged from seven seconds to 12 seconds, with an average of 10.39 seconds. (Including the outliers, the range was from seven seconds to 20 seconds, with an average of 10.67 seconds). Total time to complete the import process ranged from eight seconds to 24 seconds, with an average of 12.52 seconds. (Including the outliers, the range is from eight seconds to 89 seconds, with an average of 14.79 seconds).

<u>Workspace Loads</u>: (33 tests). There were 33 attempted loads of skull models from previous imports into the workspace. Three images failed: one as a result of not checking the "dry" skull box; one produced only the top half of the skull for unknown reasons; and one failed due to a CT scan error (the teeth were cut off) in the original scan of a test skull.

<u>Reconstructions:</u> (36 tests). There were 36 attempted reconstructions (34 completed the process and two failed). Total time to complete a reconstruction ranged from 3 hours 6 minutes to 3 hours 20 minutes, with an average of 3 hours 15 minutes. (Including the outliers, the range was from 55 minutes to 3 hours 22 minutes, with an average of 3 hours 11 minutes).

Excluding the two outliers, there were 14 image failures; 12 positive results (increasing the remaining sample size from 10 to 12); and six cases identified as "maybes" pending testing with Beta 4.0.

Testing Results by Category (Regular, Dental Noise, Dowel/Alignment, CT Errors)

Regular Sample Testing (11 Subjects, 26 tests)
Ten UTK subjects with no previously identified problems were retested. The GE Wet Skull was included in this category as a control sample. A total of 26 tests was conducted (12 imports and 14 reconstructions).

<u>Import Skulls</u>: (12 tests). All completed the import process. DICOM import time ranged from 0 seconds to 1 second with an average of 0.6 seconds. (Including the two outliers, the range was from 0 seconds to 69 seconds, with an average of 6.33 seconds.) Creation of bone models (bone.vtk) ranged from 9 seconds to 11 seconds, with an average of 9.9 seconds. (Including the two outliers, the range was from 9 seconds to 20 seconds, with an average of 10.75 seconds.) Complete import time ranged from 10 seconds to 11 seconds, with an average of 10.5 seconds. (Including the two outliers, the range was from 10 seconds to 89 seconds, with an average of 17.08 seconds.) There were no image failures.

<u>Reconstructions:</u> (14 tests). All completed the reconstruction process. Total time ranged from 3 hours 8 minutes to 3 hours 20 minutes, with an average of 3 hours 14 minutes. (Including the two outliers, the range was from 55 minutes to 3 hours 22

minutes, with an average of 3 hours 3 minutes.) There were five image failures, and two samples eliminated (15-01D and 28-01D). Reasons for the failures of two cases (15-01D and 28-01D) could not be ascertained. DICOM data from 15-01D and 28-01D were sent to the program developers for evaluation.

Dental Noise Testing (10 Subjects, 22 tests)
Ten subjects previously identified as having dental noise problems due to teeth with metal artifacts were retested. A total of 22 tests was conducted (11 imports and 11 reconstructions).

Import Skulls: (11 tests). All completed the import process. DICOM import time ranged from 0 seconds to 13 seconds with an average of 3.82 seconds. Creation of bone models (bone.vtk) ranged from 7 seconds to 12 seconds with an average of 10.36 seconds. Complete import time ranged from 8 seconds to 24 seconds with an average of 14.18 seconds. There were two image failures.

Reconstructions: (11 tests). Two subjects failed the reconstruction process and nine completed. Total time ranged to complete a reconstruction 3 hours 10 minutes to 3 hours 19 minutes, with an average of 3 hours 15 minutes. There were no image failures. It was anticipated that all would fail; however three cases (52-03D, 35-02D, 08-02D) produced successful approximations and were returned to the sample for further testing. Six additional approximations were identified as potentially usable pending future testing with Beta 4.0 (01-02D, 31-03D, 49-03D, 05-98D, 18-93D, and 18-03D).

Dowel/Alignment Testing (7 subjects, 14 tests)
Seven subjects were previously identified as having problems due to an exposed wooden dowel, poor alignment of the mandible due to dowel slippage, or both. A total of 14 tests was conducted (seven imports and seven reconstructions).

Import Skulls: (7 tests). All completed the import process. DICOM import time ranged from 0 seconds to 1 second with an average of 1 second. Creation of bone models (bone.vtk) ranged from 9 seconds to 12 seconds with an average of 10.71 seconds. Total time to complete the import process ranged from 10 seconds to 13 seconds with an average of 11.57 seconds. There were no image failures.

Reconstructions: (7 tests). All completed the reconstruction process. It was anticipated that all would fail the facial approximation (images) and six out of seven did fail. Surprisingly, one case did not fail (31-01D) and it was returned to the sample for further testing.

CT Scan Error Testing (3 subjects, 9 tests)
Three subjects previously identified as having CT scan error problems were retested. A total of nine tests was conducted (five imports and four reconstructions).

Import Skulls: (5 tests). Two subjects failed the import process. DICOM import time ranged from 0 seconds to 8 seconds with an average of 4 seconds. Creation of bone models (bone.vtk) ranged from 11 seconds to 12 seconds with an average of 11.33 seconds. Complete import time ranged from 12 seconds to 19 seconds with an average of 15.33 seconds. There was one image failure.

Reconstructions: (4 tests). All completed the reconstruction process. Total time to complete a reconstruction ranged from 3 hours 11 minutes to 3 hours 17 minutes, with an average of 3 hours 14 minutes. There were three image failures. It was anticipated that all would fail; however one case produced a successful approximation and was returned to the sample for further testing (42-01D).

Validation IV (Beta Version 4.0)

(November 10, 2005 to March 14, 2006). Installation of this version initially failed to install. It was subsequently determined that there was insufficient space on the hard drive for the installation to complete. There was no provision for archiving or deleting previous testing data within the program. All testing data were manually archived to DVDs. A second backup was made on an external hard drive. All testing data (skulls and reconstructions) were manually deleted from ReFace. Beta Version 4.0 successfully installed on February 14, 2006. Limited testing was conducted as a new version was expected to be installed on March 15, 2006. Testing of the new user interface modifications and reconstruction of one Bone Clones® replica skull was done.

Testing with HYTEC FCT-3200 CT Flash Industrial CT Scanner (HYTEC) and Bone Clones® Replica Skulls

Import Skulls. A high resolution scan (estimated at 0.2 mm slice thickness) from an industrial CT machine (HYTEC) was done on a Bone Clones® replica skull (White Male Model BC-107). The output file was in a proprietary format and was sent to the manufacturer for conversion to DICOM data file format for import into ReFace. The converted DICOM files from the White male replica skull (BC-107) were imported into ReFace. Total import time was 206 seconds, and the skull model completed with excellent detail.

Reconstruction: Total reconstruction time was 4 hours and 20 minutes. The skin image (skin.vtk) moved very slowly in the Current Estimate window. Extrema for Effect 1 appeared to be identical to the Current Estimate. Attempts to view Extrema for Effect 2 locked the program forcing a reboot.

Summary and Conclusions

A DICOM converter was supplied by HYTEC for conversion of additional high resolution scans of other Bone Clones® replica skulls. The converter failed to function, and no additional scans could be imported into ReFace. The error message stated: “The source data does not match the requested X and Y image dimensions.” Data files requested and sent to HYTEC for evaluation were: config.ini, header file, OutofOrderComposit.ini, and the print screen error messages. HYTEC has been unable to resolve the DICOM converter problem as of this date.

RE/FACE Validation V (Beta Version 5.0)

Validation testing began on March 15, 2006 and ended on June 14, 2006. Beta 5.0 introduced “canonical” males and females for each ancestral group to represent the “average” of all the faces in a given population (Peter Tu, personal communication). Additionally, the entire program was refocused on “unique skulls” that would be imported once, and then multiple reconstructions could be made from the same skull. Together with a reduction in file size of the bone and skin models, the archival problem from previous versions was resolved.

After nine reconstructions were completed and compared, it was realized that all looked like their respective canonical model. Additionally, each skin.vtk for every subject in the database (218 subjects) also looked like its respective canonical model. The result was that most variation had been eliminated and all reconstructions looked the same. No further reconstructions were attempted until this error could be corrected. All sample skulls were imported into the program as the import process was not impaired.

Corrections, changes, and additions to the program were tested. Corrections verified to be working in this version were the age-down selection, text-log file, missing-mandible, and dowel-removal reconstructions. The ability to delete questioned skulls from the Admin menu was only partially functional. New problems discovered are Admin limitation in viewing reconstructions, screen resolution issues, sort selection for height and weight, and nasal bones extending outside the skin.vtk resulting in the nose being incorrectly located on the face.

Beta 5.0: Overall Testing Results (54 Subjects, 82 Tests)
A total of 82 tests was performed on 54 subject skulls after the update was applied. There were 73 import attempts (two failed); 71 attempts to open skull models (bone.vtk) in the workspace (31 image failures); and nine attempted facial approximations (no failures).

Import Skulls: (73 tests). New Imports of all 53 UTK subject skulls were done under new case numbers. The GE wet skull was also tested. The first sample of CT scans was done at 5.0 mm slice thickness. There were 38 scans at 2.5 mm slice thickness in the second sample of CT skull scans. As slice thickness had been found to be correlated to import time, two skulls were also scanned at 1.25 mm and 0.625 mm slice thickness for comparison.

> *Outliers:*
> The GE wet skull (2.5 mm) was inadvertently tested with the head tray series and the import numbers were much higher than the series without the head tray. The two 0.625 scans were also much higher than the remainder of the import times. These three subject tests are included as outliers in the summaries.

Averages without these three outliers are given first and then averages including these three outliers. There were 73 import attempts (71 completed and two failed the import process). Total time to copy the DICOM data ranged from 0 seconds to 106 seconds with an average of 5.87 seconds for all subjects; excluding the three outliers, the import time ranged from 0 seconds to 15 seconds with an average of 4.22 seconds.

Creation of the bone model image (bone.vtk) ranged from 10 seconds to 68 seconds, with an average of 16.76 seconds; excluding the three outliers, the import time ranged from 10 seconds to 29 seconds with an average of 15.51 seconds. Total time to complete the import process ranged from 10 seconds to 126 seconds, with an average of 22.63 seconds; excluding the three outliers, the import time ranged from 10 seconds to 34 seconds with an average of 19.74 seconds.

Import Failures: (Two failures)
Case 43-03D:
Failures were due to DICOM "scout" images in the original DICOM data. It was necessary to view the DICOM data in a CT reader to see the full front and side (scout) images as the first and second files. These two files were moved to a new folder under a new series. Reimporting the DICOM data starting with the third file image was successful.

GE3 (SE2) (outlier)
GE3 (SE3) was selected for a new import. This series contained the head tray (import completed but image failed). Import time was 126 seconds, and was included as an outlier in the import times.

Workspace Loads: There were 71 attempted loads of skull models (bone.vtk) from previous imports into the workspace. There were 32 image failures.

1 = Metal plate on chin with small spikes in the bone model
17 = Spikes ranging from slight to severe from dental noise
1 = Bone model tilted in workspace
1 = Head tray visible in skull model (wrong series selected for GE3)
1 = Small bone anomalies under skull model (corrected in Beta 5.1)
6 = Dowel images present in skull model (corrected in Beta 5.1)
5 = CT errors (from severe to slight)

Reconstructions (9 Tests) There were nine attempted reconstructions (9 completed the process and 0 failed). Total time to complete a reconstruction ranged from 0 hours 3 minutes to 2 hours 45 minutes, with an average of 1 hour 27 minutes. (Three minutes was for a second reconstruction with only minor alterations of the demographic data.) Age down selection of 30/49 or 30/50 was approximately 1 hour and 20 minutes; and the full database mappings (49/49 or 50/50) exceeded two hours but were less than three hours. There were no image failures.

Testing Results by Category (Regular, Dental Noise, Dowel/Alignment, CT Errors)

Regular Sample Testing (32 Subjects, 50 tests)
Thirty UTK subjects with no previously identified problems were retested. The GE wet skull was included in this category as a control sample. One UTK skull with dowel images manually removed was included as a test of the dowel removal protocol. A total of 50 tests was conducted (41 imports and 9 reconstructions).

Import Skulls: (41 tests). There were 41 total tests (39 completed the import process and two failed). Import failures resulted from DICOM "scout" images being included in the DICOM directory. DICOM import time ranged from 0 seconds to 19 seconds, with an average of 3.87 seconds. (Excluding two 0.625 outliers, import time ranged from 0 seconds to 15 seconds with an average of 3.43 seconds.) Creation of bone models (bone.vtk) ranged from 10 seconds to 68 seconds, with an average of 19.18 seconds. (Excluding the two outliers, the range was from 10 seconds to 31 seconds, with an average of 17.11 seconds.) Complete import time ranged from 10 seconds to 87 seconds, with an average of 23.05 seconds. (Excluding the two outliers, the range was from 10 seconds to 34 seconds, with an average of 20.54 seconds.) There were no image failures.

Reconstructions: (9 tests) All completed the reconstruction process. Total time ranged from three minutes to 2 hours 45 minutes, with an average of 1 hour 17 minutes. There were no image failures in that each reconstruction produced a complete face; however, all nine reconstructions were identical to their respective canonical model (CAU females looked like the CAU female canonical model and to each other; CAU males looked like the CAU male canonical model and to each other). No further reconstructions were done pending the solution of this problem.

Dental Noise Testing (17 Subjects, 17 tests)
Seventeen subjects were tested. A total of 17 tests was conducted (17 imports and no reconstructions).

Import Skulls: 17 tests. All completed the import process. DICOM import time ranged from 0 seconds to 13 seconds with an average of 2.06 seconds. Creation of bone models (bone.vtk) ranged from 10 seconds to 21 seconds with an average of 14.53 seconds. Complete import time ranged from 10 seconds to 24 seconds with an average of 16.59 seconds. As dental spikes are known to distort the cheek and lip area, all were listed as image failures.

Reconstructions: There were no reconstructions attempted with the above 17 subjects.

Summary of Dental Noise Testing
Dental spikes, even moderate ones, appear to distort the cheek and lip area in reconstructions. All are listed as image failures in this report as reconstructions cannot be attempted until the next Beta version has been delivered. The only known way at this time to determine if the dental noise has made a difference in the cheek and lip area is to run reconstructions and compare them with antemortem photographs. The researcher developed a protocol to remove some of the dental noise from the original CT DICOM scans; it will be tested in Beta Version 5.1. As most forensic skulls will have some type of metal dental artifacts, this was an area that had to be addressed before ReFace could become fully functional.

CT Scan Error Testing (4 Subjects, 5 Tests)
Four subjects were identified as having CT scan error problems. A total of five tests was conducted (five imports and no reconstructions).

Import Skulls: (5 tests). All completed the import process. DICOM import time ranged from 1 second to 15 seconds with an average of 10.6 seconds. Creation of bone models (bone.vtk) ranged from 12 seconds to 20 seconds with an average of 13.6 seconds. Complete import time ranged from 21 seconds to 27 seconds with an average of 24.20 seconds. All images failed.

Dowel/Alignment/Misc Testing (10 Subjects, 10 tests)
Ten subjects were identified as having problems due to: exposed wooden dowel (6); poor alignment of the mandible due to dowel slippage (1); metal plate on chin (1); head tray around head (1); and unknown bone anomalies (1). A total of ten tests was conducted (10 imports and no reconstructions). A test for dowel removal was done and included in the Regular Sample (14-92D, dowel removed). The bone model showed no sign of the previous dowel image and a reconstruction was completed.

<u>Import Skulls</u>: (10 tests). All completed the import process. DICOM import time ranged from 0 seconds to 106 seconds with an average of 17.8 seconds; excluding the GE3 wet skull outlier (with head tray), the import time range was from 0 seconds to 12 seconds with an average of 8 seconds. Creation of bone models (bone.vtk) ranged from 10 seconds to 20 seconds with an average of 12.7 seconds; excluding the GE3 outlier, the time range was 10 seconds to 20 seconds with an average of 11.89 seconds. Total time to complete the import process ranged from 10 seconds to 126 seconds with an average of 30.5 seconds; excluding the GE3 outlier, the time range was from 10 seconds to 25 seconds with an average of 19.89 seconds. All were image failures.

<u>Reconstructions</u>: There were no reconstructions on the above subjects.

RE/FACE Validation V (Beta Versions 5.1/5.2)

Beta Version 5.1 was installed on June 14, 2006 and testing ended on November 7, 2006. A program error prevented African ancestry males and females from completing the reconstruction process. This problem was corrected by a patch installed on November 8, 2006 (as Beta Version 5.2.), and testing ended on November 30, 2006.

<u>Major Changes in Beta Version 5.1</u>**:**
Beta 5.1 corrected the image problem discovered in Beta 5.0 that caused all the faces to look the same. New skin.vtk files were provided for all database subjects to “fix the perception issue caused by extraneous data” (Peter Tu, personal communication). All new database skin.vtk files were reviewed and all UTK skulls were retested. The CAU male and female reconstructions no longer look like the canonical modes. The CAU males also look different from each other; however, there is strong similarity between the CAU females.

New Developments
<u>Extraneous “Bone” Fragments in Skull Models</u>
Extraneous fragments from articulation and packaging materials appeared in the CT DICOM data and were perceived as “bone” fragments by ReFace. The program attempted to incorporate this information into the skull model (bone.vtk), resulting in deformed current estimates, extrema, and image failures. The program developers were successful in manually removing the “bone” anomalies as well as dowel images from the skull models. The original bone.vtk files were renamed and the corrected versions were copied into the model directory. The program used the corrected bone.vtk files for facial approximations. This manual method worked, but an automated method was needed to removed extraneous material from the bone models.

Beta 5.1/5.2 Overall Testing Results (55 Subjects, 100 Tests)
A total of 100 tests was performed on 55 subject skulls after the Beta 5.1 and Beta 5.2 updates were applied. There were eight new imports of CT data from unknown skulls entered into the program (one failed); seven attempts to open the skull model (bone.vtk) in the workspace (one image failed); and there were 92 attempted facial approximations (five failed to complete process; 34 images failed).

Import Skulls: (8 tests). Imports of all 53 UTK subject skulls and the Bone Clones® replica skull (BC-107) were previously completed in Beta 5.0. Eight new imports were attempted after installation of Beta 5.1 (seven imports completed and one failed).

> Subject 29-04D: (3 tests): Small anomalies were found under the bone model which were also visible on the original CT DICOM data. These anomalies were later determined to be capsule fragments from the packing material used in the second set of CT scans. The anomalies were later manually removed from the bone models (bone.vtk) by the program developers.
>
> Subject 12-04D: (4 tests): Metal dental artifacts caused spikes on the bone.vtk resulting in reconstruction image failure in Beta 5.0. Three tests failed in the development of a new protocol to remove the dental spikes prior to import. (This protocol is discussed in Appendix D). The fourth test successfully removed the dental spikes from the bone.vtk and produced a normal reconstruction.
>
> BC-107: (1 test): Bone Clones® BC-107 was scanned with an industrial CT scanner at an estimated 0.2 mm slice thickness. The import completed but the image failed due to an unknown program error. The bone.vtk was produced but did not appear in the workspace.

Import Testing Results:
The UTK subjects were 2.5 mm slice thickness; the replica skull (BC-107) was scanned at 0.2 mm slice thickness. There were eight import attempts (seven completed and one failed). Total time to copy the DICOM data for the UTK subjects was one second; creation of the bone model image (bone.vtk) ranged from 15 seconds to 20 seconds with an average of 16.83 seconds. Total time to complete the import process ranged from 16 seconds to 21 seconds with an average of 17.83 seconds. BC-107 time to copy the DICOM data was 23 seconds; creation of the bone model image (bone.vtk) was 7 seconds; and total import time was 30 seconds.

Workspace Loads: (7 tests). There were seven attempted loads of skull models from previous imports into the workspace. One failed the loading process (BC-107); one was successful (12-04D new protocol); and there were five image failures (three bone anomalies and two dental spikes).

Reconstructions: (92 tests). There were 92 attempted reconstructions (87 completed the process and five failed). Total time to complete a reconstruction varied depending on the number of mappings from the database. Reconstruction time ranged from 0 hours 25 minutes (10 out of 10 mappings) to 2 hours 18 minutes (49 out of 49 mappings), with an average of 1 hour 23 minutes.

There were no UTK skulls of Americans of Asian ancestry available for testing. Asian replica skulls were scanned with an industrial scanner for testing purposes only, but the data could not be converted to DICOM for import into ReFace. For testing purposes only, imports and facial approximations were done on a White female subject (28-01D) and a White male subject (10-02D) to test the Asian male and female mappings. Both completed the reconstruction process but were image failures (exposed nasal bones). Table C.1 summarizes the 92 attempted reconstructions.

Reconstruction Image Failures
Of the 87 completed facial reconstructions, 34 were image failures. Errors in articulation of the mandible (1), errors in CT scans (6), and anomalies from packing materials (5) accounted for 12 of the failures. Dental spikes from metal dental artifacts resulted in seven image failures. The remaining image failures were due to incorrect placement of the nose due to nasal bones extended outside the skin image (11), both dental spikes and extended nasal bones (3), and an unknown program error (1).

Testing Results by Category
(Regular, CT Errors, Dental Noise, Dowel/Alignment/Bone Anomalies/Misc, Program Errors)

Regular Sample Testing (22 Subjects, 29 tests)
There were 29 tests performed on 22 subject skulls (no new imports and 29 reconstructions). Eight tests were on previous image-failed reconstructions from Beta Version 5.0 that were reopened in Beta 5.1. Seven correctly opened with a different face than the canonical face and one reconstruction failed to open.

Import Skulls: There were no new imports in this category.

Reconstructions: (29 tests) Seven previous reconstructions opened and one previous reconstruction failed to open. There were 21 new reconstructions and all completed the reconstruction process. Total time ranged from 37 minutes (15/15 mappings) to 2 hours 18 minutes (49/49 mappings), with an average of 1 hour 24 minutes. There were no image failures. Improvements in the program reduced the average reconstruction time from more than three hours to less than two hours.

CT Scan Error Testing (4 Subjects, 5 Tests)
Four subjects were identified as having CT scan error problems from previous testing. Five reconstructions completed the reconstruction process. Reconstruction time ranged from 1 hour 21 minutes to 1 hour 35 minutes with an average of 1 hour 25.8 minutes. All were image failures. All CT scan error subjects were with White males with chins cut off in the CT scan. It was critical that the entire skull be contained in the CT scan.

Dental Noise Testing (16 Subjects, 21 tests)
Seventeen subjects were tested. A total of 21 tests was conducted (four imports and 17 reconstructions).

Table C.1: Summary of Reconstruction Tests (Beta 5.1 and 5.2).

Database Total	Database Population	Total Tested	Average Time	Tests
N=10	Black Females	(10/10)	0 H 25 M	1
N=15	Black Males	(15/15)	0 H 37 M	4
N=50	White Females	(30/50)	1 H 39 M	25
	White Females	(50/50)	2 H 09 M	2
N=49	White Males	(30/49)	1 H 26 M	43
	White Males	(49/49)	2 H 18 M	1
N=47	Asian Male	(47/47)	2 H 08 M	1
N=47	Asian Female	(47/47)	1 H 49 M	1

Import Skulls: (4 tests). Three completed the import process and one failed. DICOM import time for each of the three imports was one second. Creation of bone models (bone.vtk) ranged from 15 seconds to 16 seconds, with an average of 15.33 seconds. Complete import time ranged from 16 seconds to 17 seconds, with an average of 16.33 seconds. There were two image failures due to dental spikes. The third test completed successfully after using the researcher's protocol for the removal of dental spikes from the bone.vtk image prior to reconstruction. This protocol is discussed in Appendix D.

Reconstructions: (17 tests). There were 17 reconstructions attempted and all completed the reconstruction process. Total time ranged from 25 minutes (10/10 mappings) to 1 hour 35 minutes (30/49 mappings), with an average of 1 hour 17.11 minutes. There were ten image failures: dental spikes (7), and dental spikes and extending nasal bones (3).

Conclusions (Dental Noise Testing)
The only known way at this time to determine if dental noise has made a difference in the cheek and lip area is to run reconstructions and compare them with antemortem photographs. Of the 53 UTK sample skulls with antemortem photographs, 14 were eliminated from the testing due to dental spikes caused by metal dental artifacts. Seven UTK subjects were reconstruction image failures due to dental spikes alone; three failed due to dental spikes and nasal bone projections. Another four subjects with slight dental spikes were placed in a test category and the reconstructions compared to the lip and cheek areas of antemortem photographs of the subjects.

Dowel/Alignment/Bone Anomalies/Misc Testing (14 Subjects, 28 tests)
Fourteen subjects previously identified with problems in Beta 5.0 were retested in Beta 5.1. The bone.vtks of those with visible dowels and bone anomalies had been manually corrected by program developers. A total of 28 tests was conducted.

Import Skulls: (3 tests). All completed the import process. DICOM import time for all three subjects was one second. Creation of bone models (bone.vtk) ranged from 17 seconds to 20 seconds with an average of 18.33 seconds. Total time to complete the import process ranged from 19 seconds to 21 seconds with an average of 19.33 seconds. All were image failures (bone anomalies on bone.vtks).

Reconstructions: (25 tests). There were 25 reconstructions attempted and all completed the reconstruction process. Total time ranged from 1 hour 12 minutes (30/50 mappings) to 2 hours 1 minute (50/50 mappings), with an average of 1 hour 25.62 minutes. There were eight image failures: improper alignment of mandible (1); bone anomalies (4); exposed nasal bones (3).

Conclusions (Exposed Dowels and Bone Anomalies)
Exposed Dowels: Dowel images appearing on bone.vtk images were successfully removed by program developers using a manual process. This enabled five previously excluded UTK subject to be returned to active validation status. The problem was not found in the second sample of UTK skulls as they were articulated with dowels that did not extend outside the skull.

“Bone” Anomalies: There were small anomalies that appeared to be bone fragments in the bone.vtk images. Most were later determined to be broken capsules from the packing materials used in the second sample. One skull from the first sample (37-02D) was found to have the top of a second CT scan. When present, such extraneous material caused image failures in the Current Estimate and Extrema windows. These anomalies were successfully removed from the original DICOM files by a protocol developed by the researcher (see Appendix D). The program developers were able to remove these anomalies from the bone.vtk images using a manual method. (However, some were still image failures due to the new problem of exposed nasal bones.)

Program Error Testing (13 Subjects, 17 tests)
Thirteen subjects were tested. A total of 17 tests was conducted (one import and 16 reconstructions).

Import Skulls: (1 test). Subject BC-107 was the only import test in this category. BC-107 (White male) is a replica skull from Bone Clones® which was scanned at an estimated 0.2 mm slice thickness with an industrial CT scanner (HYTEK). Time to copy the DICOM data was 23 seconds, create the bone.vtk was 7 seconds, and total import time was 30 seconds. The bone.vtk failed to load in the program workspace for unknown reasons.

Reconstructions: (16 tests). Twelve completed the reconstruction process and four failed. One failure was caused by an unknown program error, and three failures were due to a program error in Beta 5.1. This new error caused reconstructions of Black males and Black females to abort after the final database mapping. A program correction was installed as Beta Version 5.2. Total time ranged from 30 minutes (15/15 mappings) to 2 hours and 8 minutes (47/47 mappings), with an average of 1 hour 25.63 minutes.

There were ten image failures: one image failed to cover the chin (86-18F) and nine had exposed nasal bones with incorrect nose positions. No correlations were found that might explain these failures. Three subjects with dental spikes also have the “nose” problem.

Unresolved Issues as of ReFace Beta Version 5.2
Programming areas for correction include screen resolution inconsistency, an option to delete bone models, feature selection sorting, Admin(viewing limitations, complete installation portable media, update of help files, and a user tool to remove extraneous data.

Appendix D

Phase 1: Protocol Development for Articulation, Packaging, and Transport of Skulls for CT Scanning

Appendix D
Phase 1: Protocol Development for Articulation, Packaging, and Transport of Skulls for CT Scanning

Articulation of the Mandible and Cranium

Traditional Methods of Articulation
In skeletal remains, the mandible (lower jaw) and the cranium (remainder of the skull) is not articulated as the soft tissue that holds them together has decomposed. For proper alignment, the mandible and cranium must be articulated with some kind of material to serve as a “spacer” for the missing soft tissue. If the skull is edentulous (without teeth), the soft tissue (gums) between the dentures and the bone underneath also needs to be replicated.

<u>Edentulous Skulls</u>
In Taylor (2001), Betty Pat Gatliff outlines the proper articulation of skulls for traditional clay facial approximations. She tested the articulation method illustrated in *Grant's Anatomy* (1983), by inserting a wooden dowel or pencil in skulls with full dentition, and found the method did indicate the proper alignment (personal communication, 2006). The Initial scans of 30 UTK skulls used this method with the wooden dowel extending outside the skull. However, it was found that the prototype interpreted this material as “bone” and attempted to add it to the skull model. (The second set of CT scans used dowels for articulation, but they were cut to the exact size so none of the wood would extend outside the skull.)

Articulation of the skull has traditionally used epoxy (Duco® Cement) and cotton balls as a spacer for the TMJ. The first set of CT scans used only materials known to be invisible for the articulation of skulls that did not require a dowel. As cotton was known to be interpreted as bone in a CT scan and the effect of the epoxy was unknown, the TMJ was articulated with folded squares of bubble wrap and foam wrap to hold the mandible and cranium together. Problems were encountered with the import of the first set of skulls due to the “space” left by the materials. When the program encountered the air pocket, it assumed there was no more bone to follow and created only the top half of the skull. Testing of other materials was done prior to the second set of scans.

Articulation Methods for CT Scanning
No standardized protocol was found in the literature for articulation of the test skulls in preparation for CT scans. Previous development testing to validate the prototype was done using a leave-one-out method: simulated sample skulls for testing were obtained by removing the skin layer and leaving the bone layer from a CT scan of a living subject from the ReFace database. This articulated and dense (wet) skull was used as an “unknown” skull for testing.

From this simulated skull, the prototype software created a three-dimensional skull model. This model was perfectly articulated at the TMJ as the soft tissue at the articulation point was still present. Articulated cervical vertebrae provided the simulated skull with an approximate Frankfurt-Horizontal position. The living bone of the simulated skull was very dense (wet). The slice thickness of the CT DICOM data was determined by the original scan of the living person; most of the living subjects in the database were scanned at a slice thickness of either 5.0 mm or 2.5 mm .

In contrast, the UTK testing skulls were not articulated, less dense (dry). They had to be physically stabilized for transport to protect the skull, and then positioned horizontally (looking up) for the actual CT scan. Some materials were known to be visible in CT scans (metal), while others were known to be *invisible* (bubble wrap, foam rubber), but the impact of most materials was unknown. Before the UTK testing skulls could be scanned, methods of articulation, packaging, and positioning had to be developed. Technical specifications for the CT machines and the optimal slice thickness for dry skulls were unknown, as dry skulls had not been tested.

Testing of Articulation and Packing Materials

Testing Skulls

Four replica skulls used for testing were obtained from Bone Clones® Osteological Reproductions: Modern Human Series (*Homo sapiens*).

1. African Male (Skull BC-110)
2. Asian Male (Skull BC-016)
3. Asian Female (Skull BC-149)
4. White Male (Skull BC-107)

Articulation Materials

A variety of materials were tested for use as articulation materials to temporarily "glue" the mandible and cranium together and to replicate missing soft tissue. Materials tested for articulation include bubble wrap, thin foam wrap, masking tape, paper adhesive tape, cotton gauze pads, cotton balls, paper adhesive tape, rubber bands, masking tape, and Duco® Cement. Three materials tested for skull articulation (all from Trevco (San Marcos, CA 92069) were designed for use by museums to anchor items to a surface area: Crystalline Clear Museum Wax © (a blended microcrystalline wax); Clear Museum Gel© (a removable, reusable, translucent gel); and The Museum Putty© (also for wood, tile, and walls).

<u>Results</u>*:*

Bubble wrap, thin foam wrap, masking tape, and paper adhesive tape were invisible to the CT machine and to the program. Thin rubber bands were not visible in the 5.0 mm CT scans, but were visible on the industrial CT machine scans (estimated to be 0.2 mm). These materials were used in the first UTK sample and enabled rapid articulation of the skulls and minimal time to disarticulate them. It was unknown if (1) the rubber bands would be visible on the 2.5 mm, 1.25 mm, and 0.625 mm scans planned for a second sample, and (2) whether the visibility would impair the bone model created by the program.

Cotton (gauze pads and cotton balls) together with Duco® Cement were tested on the Bone Clones® replica skulls. Any part of the cotton that extended outside the skull was interpreted as bone by ReFace, but was not visible if kept inside the skull. The continuity between mandible and cranium by the materials eliminated the air pocket created by materials found to be invisible to the CT machine.

Crystalline Clear Museum Wax© , Clear Museum Gel©, and The Museum Putty© all worked on the Bone Clones® replica skulls, and all three were tested on the second set of CT scans of UTK human skulls. However, when the museum gel was used on actual human skulls, the material melted during the CT scan process and formed a thin, sticky layer of "goo" over the bone surface that was difficult to remove. The museum putty maintained its consistency and was interpreted as "bone" by the program (the same as cotton). The museum wax also maintained its consistency, but was more difficult to clean from the bone than the museum putty.

Dowel Materials
Wooden dowels used for articulation of edentulous skulls that extended outside the skull area were interpreted as bone. Eight different types of dowel replacements were tested with the replica skulls. The samples were inserted in a block of hard foam and CT scanned as a single unit.

Replacement Dowels Tested
1. Regular wooden dowel for comparison
2. Single candy sucker paper stick (3/16 diameter)
3. Candy sucker paper sticks (4) with rubber bands on ends (2/16 diameter)
4. Rolled up copy paper with ends secured by rubber bands
5. Small plastic tube (hollow) (1/8 inch diameter)
6. Larger plastic tube (hollow) (approx. 1/4 inch diameter)
7. Regular wooden dowel wrapped in thin layer of foam rubber secured on the ends with masking tape
8. Thick opaque hollow plastic tube

Results:
All substitutes were visible on the CT viewing screen and densities were equal to or greater than that of the wooden dowel. This scan was done on an industrial machine, and the settings did not easily translate to those of CT scanners used for humans. Prior to elimination as testing materials, these same samples should be scanned again with known parameters.

Packaging Materials
A method to prevent movement of the skulls during the CT scan process was needed to keep the dowels in place and prevent movement of the skull. Numerous packaging materials were tested with the replica skulls to develop a stabilization method. Materials were tested with and without being contained in cardboard boxes.

Materials Tested
1. Soft foam rubber
2. Medium Styrofoam™
3. Hard Styrofoam™
4. Medium plastic foam with hard white plastic coating
5. Soft-Pak® (gray plastic covering over soft foam rubber).
6. Insta-Pak Quick® RT

Results:
Skulls contained in cardboard boxes during the CT scan were more stable than those balanced on models created in foam and wrapped in bubble wrap. Boxes could be sealed by tape, which protected the skulls and minimized time during transport and scanning. Previous problems caused by the conveyor movement and vibration were eliminated.

For a single skull, the easiest method was to create a support base of foam (soft or hard) cut to exactly fit the skull. The base was then covered in bubble wrap, which provided and invisible layer between the skull and the packaging materials. This method provided a secure, protective, and stable base for CT scans, but the individualized construction was time consuming. Standardizing this method to accommodate skulls of different sizes, or a large volume of skulls at the same time, did not seem feasible. A replicable, fast, and easy method was needed for the articulation of the second UTK sample of 35 skulls.

One of the materials tested and confirmed to be invisible to the CT machine was an expandable foram product, Soft-Pak®, but it required the purchase of a machine for bag inflation. The researcher became aware of a similar foam packing material, Insta-Pak Quick® RT, which inflated by chemical reaction and required no additional equipment or machinery. As the packaging of human skulls was not a traditional use of this material, the product requirements to do so were unknown. The manufacturer provided sample bags of different sizes were provided by the manufacturer for testing purposes. A variety of configurations were tested to determine the best method for packaging human skulls.

Protocols

Articulation of Edentulous Skulls
Wooden dowels cut to the *exact* length of each specific skull and glued into place do not interfere with the reconstruction process. Traditional use of Duco® Cement was not sufficient to maintain proper articulation when the skull was placed in a horizontal position without additional support (such as packaged in a box). Movement and vibration of the conveyor belt of the CT machine caused the skulls to move out of position; dowels vibrated out on one or both sides and had to be reset several times before the scan could be completed. Wooden dowels, though not visible in the viewing screen during the CT scans, did appear in the bone models of the software after import of the CT DICOM data.

Packaging
The optimum packaging was found to be two Insta-Pak Quick® RT Size 10 bags, contained in a 12 x 10 x 10 cardboard box. The bags are normally used in pairs (top and bottom); one bag is placed on the bottom (creating a base around the item) and the other bag on the top (inflated and contained by closing the box). This method was not suitable for skulls as the pressure from inflating the top bag over the “face” could have damaged the skull.

For skull packaging, optimum placement was found to be on the top and bottom surfaces of the horizontal skull, not the top and bottom of the box. One bag was inflated and placed at one long end of the box. One edge of the bag was centered in the bottom of the box and the other edge was draped over the outside top of the box. The skull was *quickly* placed in the box during inflation so that the expanding foam formed around the top of the skull. The skull was then removed and a second bag inflated and placed at the opposite long end of the box. The skull was returned to the box as the second bag inflated around the bottom (mandible) of the skull. The skull was held in place until the bag was fully inflated. This procedure created a secure, stable foam base that prevented any movement of the skull.

A caveat to those who wish to try this procedure: once the capsules are broken to mix the chemicals, the bags inflate *very* quickly and the surface area is very hot, necessitating protective gloves. Additional information about this packaging protocol can be obtained by contacting the researcher.

Skull CT Scans

CT Scan Errors

It is critical that the *entire* skull be included in the CT scan (medical diagnostic scans do not have this requirement, and often the top of the head or the tip of the chin is cut off). An incomplete CT scan, or use of incompatible packing or articulation materials, did cause the reconstruction image to fail. A specific protocol and detailed instructions should be provided to the persons responsible for CT scanning an unknown skull for import into ReFace.

DICOM Scout Images

These files are complete skull images in front or profile views which originate from resizing the CT scanning window (such as to accommodate a larger than normal skull). Import of the skull DICOM data will fail if there are “scout” images included in the same DICOM directory as the skull slices. Scout images can be identified by viewing the original DICOM data in a viewer, or by viewing file size details in a computer directory. Scout files are usually listed first in the DICOM directory and are typically double in size from the remaining slice files. Once identified, they can be removed from the directory and saved as a secondary file. There is no need to renumber the remaining files in the DICOM directory after removal of the scout files.

CT Slice Thickness

The optimum slice thickness for import was determined to be 2.5 mm. Extremely high resolution scans (anything more than .625 mm slice thickness) slowed the software down to an unacceptable speed and provided no additional quality in the facial approximation. CT scans can be transferred to portable media and imported into the software program. The prototype compiles the individual CT scan slices and creates a three-dimensional skull model that is used for facial approximation.

Removal of Dental Noise

The researcher developed a protocol of removing single DICOM slices from the original DICOM prior to import into the program. This protocol was designed for testing purposes to enable reconstructions on previously eliminated samples due to dental noise. Extraneous material above or below the skull can also be removed by this method.

The original DICOM data was viewed and the most severe flash image was identified and removed by moving the file to a different directory. Only *one* DICOM slice can be removed; more than two slices cause unacceptable alteration in the bone model image. It was expected that the computer would interpolate a value based on the two adjacent slices, or alternatively, the remaining slices would maintain their spacial coordinates. However, this does not appear to be the case. The space represented by the missing slice was deleted and the two adjacent slices become contiguous.

The remaining DICOM data files were then reimported into ReFace and a normal bone model (bone.vtk) was produced. Depending on the slice thickness, the removal of one slice should cause minimal changes in the bone model, and a normal reconstruction can be accomplished. This protocol was successfully tested on three UTK subjects for repeatability with dental spikes as well as extraneous material from packaging or scanning errors.

One subject, UTK 12-04D, had been previously removed from the validation study due to dental spikes and subsequent "puffing" out of the cheeks in the reconstruction. Application of this protocol resulted in a successful bone model and facial reconstruction. Whether severe dental spikes can be corrected by this protocol has not been tested, but preliminary testing indicates that the protocol may work for subjects with slight to moderate dental noise.

Vita

Diana Kim Moyers received a B.A. in Psychology (1992) and a B.A. in Anthropology (2002) from the University of Tennessee in Knoxville, Tennessee. Thesis research was conducted at the Counterterrorism and Forensic Science Research Unit (CFSRU), in the Laboratory Division of the Federal Bureau of Investigation in Quantico, Virginia. Diana received an M.A. in Forensic Anthropology in 2007, and is continuing validation work on the ReFace prototype in Quantico, Virginia.