2018

Reconstructing the Vocal Capabilities of Homo Heidelbergensis, a Close Human Ancestor

Austin Blake Stanley
University of Central Florida

RECONSTRUCTING THE VOCAL CAPABILITIES OF HOMO HEIDELBERGENSIS, A CLOSE HUMAN ANCESTOR

by

AUSTIN STANLEY

A thesis submitted in partial fulfillment of the requirements
for the Honors in the Major Program in Anthropology
in the College of Sciences
and in The Burnett Honors College
at the University of Central Florida
Orlando, Florida

Spring Term, 2018

Thesis Chair: Dr. John Starbuck

ABSTRACT

The discovery of 5,500 *Homo heidelbergensis* fossil specimens at the Sima de los Huesos archaeological site in Spain has opened up the opportunity for research to be conducted on the vocal capabilities of this species. Previous research has revealed that the range of vowel sounds an individual can produce, known as the vowel space, is directly affected by the dimensions of the vocal tract. The vowel spaces of two hominins, *Homo sapiens* and *Homo neanderthalensis*, have been reconstructed through previous research. However, the vowel space of *Homo heidelbergensis* has not yet been reconstructed. In this research, I aim to explore how the dimensions of the *Homo heidelbergensis* vocal tract affect the vowel space of that species. This was pursued by measuring the craniospinal dimensions of five *Homo heidelbergensis* specimens through three dimensional imaging software. When measurements were unattainable due to limitations in the fossil record, regression equations were used to predict missing measurements. By doing so, the vowel space of this species was reconstructed, and crucial information into the vocal capabilities of this close human ancestor was revealed.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Humankind has pondered about the origins of language for thousands of years. More recently, an extensive amount of research on the speech potential of *Homo neanderthalensis* (Barney 2012; Granat 2007; Steele 2013) and *Homo erectus* (MacLarnon 1999; Johansson 2011) has been completed by anthropologists. *Homo heidelbergensis* is a hominin that may have a significant phylogenetic connection to modern humans, yet direct research on the vocal capabilities of this species has been scarce. According to traditional phylogenetic models, *H. heidelbergensis* is often considered to be a direct ancestor to *Homo sapiens*. Thus, this research aims to reveal if this close human ancestor had the potential for human-like speech, as well as to assess how the anatomical characteristics involved in speech may have evolved in the human lineage.

Speech, Language, and Vocal Anatomy

Before evaluating the communicative abilities of *H. heidelbergensis*, it is first important to establish the difference between language and speech. Language is an adaptable system of symbolic vocalizations, gestures, or visual depictions that convey information and is organized with a grammatical structure. Speech on the other hand is the physical articulation of a variety of distinct and distinguishable sounds, which may be combined in an infinite assortment of combinations, and are produced by the vocal tract (Liberman, 2000). While nonhuman animals can communicate through vocal signals, the systems of calls with which they communicate are relatively static. Unlike animal communication, human language is an adaptable system that can

convey new concepts by generating new combinations of phonemes (individual units of sound), and that can assort these sequences of phonemes in an infinite number of combinations.

It is important to note that language and speech are not necessarily codependent (Liberman, 2000). Language can exist without speech. For example, sign language is a complete system of symbolic gestures, organized in a grammatical structure, that does not require any vocalizations. Speech can also exist without language. It has been suggested that glossolalia, or "speaking in tongues", represents a system of distinct vocalizations that do not carry specific meaning in an organized structure (Samarin, 1972). Parrots are also capable of producing a wide variety of distinct phonemes, but while they may be capable of connecting single words to individual concepts such as a shape or a color, this system of communication lacks the ability to combine words in a flexible structure to convey new information.

For a system of vocalizations to properly function as speech, the organism first must be capable of producing a variety of audibly distinguishable sounds. Modern human language requires a fairly complex combination of rapid articulations with the tongue and lips, vowel modifications by manipulating the jaw and tongue dorsum, as well as long exhalations of air followed by short inhalations at controlled points (MacLarnon, 1999). This high degree of articular control is important for modern human language, but is not necessarily required to this extent if the system of vocal communication remains fairly simple, such as a protolanguage that contains no rapid sequencing of vocal sounds as surmised by Bickerton (1990). Nonetheless, present day chimpanzees (and other extant non-human apes) would likely not have the physical capability to produce human-like speech, even if they had the necessary cognitive abilities (Nishimura, 2006). This is because the chimpanzee tongue is situated almost entirely in the oral

cavity, and does not significantly extend back into the pharynx. This prevents the tongue dorsum from modifying the shape of the pharynx, which is necessary for producing a variety of distinct vowel sounds *(Figure 1)*. Therefore, the evidence seems to suggest that this essential adaptation arose after the evolutionary divergence from chimpanzees.

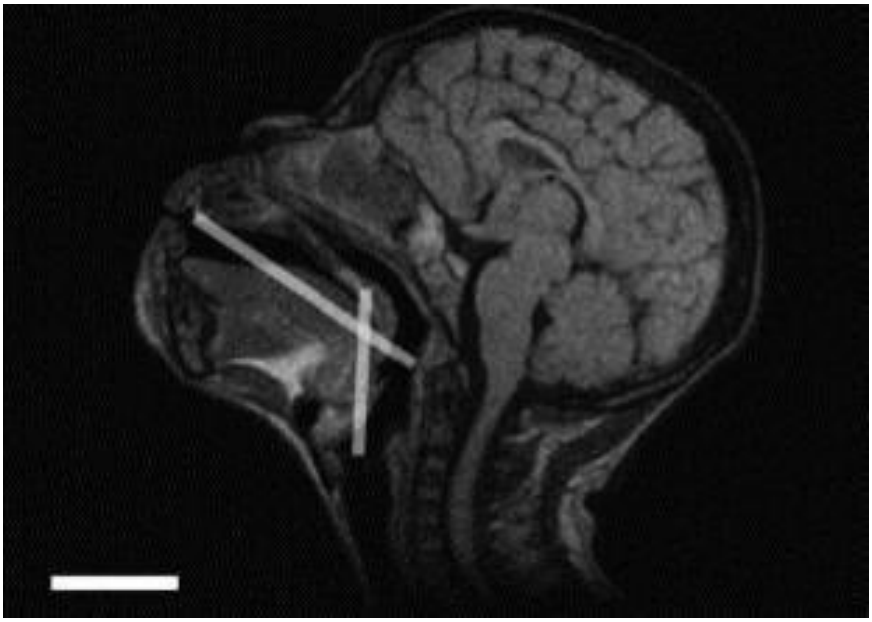


Figure 1. MR Images of a Chimpanzee Vocal Tract

SOURCE: (Nishimura et al., 2006. Figure 1, Page 3)

Vocal Tract Dimensions and the Vowel Space

Vowels are vocal sounds produced through the vibration of the vocal cords and an open vocal tract. The range of potential distinct vowel sounds an individual can produce is directly affected by the dimensions of the vocal tract. In modern humans, the vocal tract is divided into two sections: the oral cavity, also known as Vocal Tract Horizontal (VTh) and the pharyngeal cavity, known as Vocal Tract Vertical (VTv). The VTh and VTv form a "two-tube" acoustic system from soft tissue structures with a roughly 1:1 ratio (Figure 2) (Barney, 2012). Due to the

position of the tongue, humans can modify the shape of the VTh and VTv independently. This allows us to use the anterior tongue to create consonant sounds and modify vowel sounds, while using the posterior tongue as a pharyngeal wall to further modify vowel sounds.

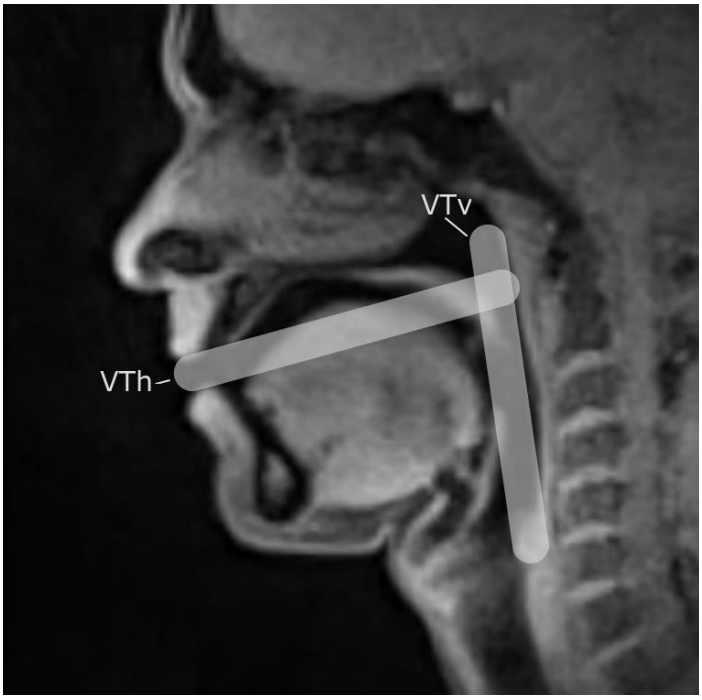


Figure 2. MR Image Depicting the VTv and VTh of a Modern Human

SOURCE: (Modified from Uecker et al., 2010. Figure 5, Page 991)

This independent manipulation of the VTh and VTv allows the vocal tract to serve as a flexible acoustic filter for sounds produced by the vocal cords. The empty space within the VTh and VTv forms a resonance chamber, or an enclosed space in which sound waves reverberate and are emitted out of an opening. The dimensions of a resonance chamber directly affect the formant frequencies that are emitted. Formant frequencies, or simply “formants”, are peaks in amplitude among the spectrum of frequencies in a sound. These formants are audibly distinguishable as different vowels based on the two lowest frequencies that are produced, F1

and F2. When all of the possible F1 and F2 frequencies that a vocal tract can produce are visualized (Figure 3), it can be observed that the potential formants of the human vocal tract create a roughly triangular range of vowels, with /a/ (as in h<u>o</u>d), /i/ (as in h<u>ee</u>d), and /u/ (as in wh<u>o</u>'d) at the corners (Barney, 2012). The maximum range of these frequencies an organism can produce is known as the vowel space.

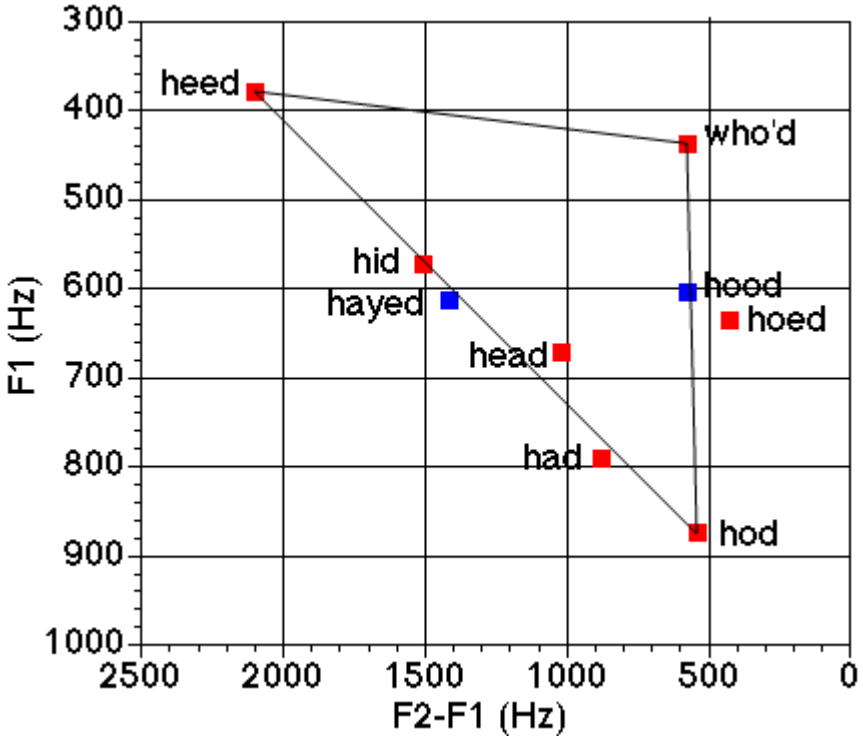


Figure 3. Representation of the Human Vowel Space

SOURCE: (USC Signal Analysis and Interpretation Laboratory)

Different combinations of F1 and F2 formants are produced depending on the orientation of the tongue in the vocal tract. Due to the "two-tube" configuration of the human vocal tract, the manipulation of the tongue's position can alter the shape of both the oral cavity and the pharyngeal cavity. In other words, the manipulation of the tongue's position can increase or

decrease the overall area of the VTh and VTv independently. This means that each vowel can be represented by a range of VTh and VTv lengths and areas that produce the formants which listeners recognize as that vowel. Thus, formants can be reconstructed by determining the length and area of a resonance chamber, such as the VTh and the VTv (Barney, 2012).

Through the work of Barney et al. (2012), the vowel space of *H. neanderthalensis* was mapped by reconstructing the dimensions of the *H. neanderthalensis* vocal tract. This was possible by measuring anatomical landmarks on the hard tissue structures of the palate, mandible, cervical spine, and hyoid, and utilizing three dimensional imaging software to create a VT model. The vowel space itself was reproduced by analysing the formants that were emitted when the *H. neanderthalensis* vocal tract was placed in three different configurations that produced the corners of the vowel space, /a/, /i/, and /u/. More details on the VT configurations for /a/, /i/, and /u/ are in the Methodology section under the *Formant Frequency Analysis* subheading.

Research Goals

This research will reconstruct the dimensions of the *H. heidelbergensis* vocal tract by utilizing three dimensional models of *H. heidelbergensis* fossil craniums and mandibles, as well as measurements of the cervical spine from Carretero et al. (1999). By obtaining linear distances from specific landmarks that represent the boundaries of the vocal tract (described below in the Methodology section), the length of the VTh and VTv will be reconstructed. Due to the correlation between VT dimensions and the vowel space of an organism, these dimensions can

be applied to the software VTCalcs to output the maximum range of F1 and F2 frequencies that could possibly be produced.

If *H. heidelbergensis* had the ability to produce a wide variety of vowel sounds, it will support the possibility that they had the capacity for human-like speech. In traditional phylogenetic reconstructions, *H. heidelbergensis* is often suggested to be a direct ancestor to modern humans (Tattersall, 2000). By evaluating the speech potential of this species, we can further narrow the range of time that human speech could have possibly begun.

In sum, it is the goal of this research to answer the following question: Was the distance between the three corners of the *Homo heidelbergensis* vowel space smaller than that of *Homo sapiens*? This question was answered by measuring the dimensions of the *Homo heidelbergensis* vocal tract, determining the formant frequencies that are produced with the /a/, /i/, and /u/ configurations within those dimensions, and estimating the outer limits of the range of vowel sounds they could have produced, in other words, the vowel space.

CHAPTER 2: LITERATURE REVIEW

The Relationship Between Speech and the Hypoglossal Canal Morphology

While we cannot directly assess the psychological capacity of *H. heidelbergensis* to communicate through language, we can observe a series of evolutionary adaptations that serve a primarily vocal purpose in the hominin lineage. Previous researchers have focused on the dimensions of the hypoglossal canal as a potential sign of articulatory capacity (Kay 1998) (Cartmill, 1998) (Jungers, 2003). The hypoglossal canal is a foramen on the occipital bone of the skull that houses the hypoglossal nerve, which is the nerve responsible for innervating the muscles of the tongue responsible for speech. It was thought that the size of the canal correlated to the degree of control an individual would have over the tongue (Kay, 1998) (Cartmill, 1998). However, the relative size of the hypoglossal canal for other hominoids such as orangutans, gorillas, siamangs, and chimpanzees, still fall within the normal range for humans, and the average relative canal size of gibbons actually exceeds that of humans (Jungers, 2003). Therefore, the size of the hypoglossal canal alone is not a reliable indicator of whether a hominoid communicates through speech.

However, even if these ranges did not overlap, the hypoglossal canal would still be a poor indicator of hypoglossal nerve size. The majority of the area within the hypoglossal canal is filled with vascular tissue and connective tissue, with the nerve itself often taking up a variable amount of space within even similar sized canals (Jungers, 2003). As such, an accurate measurement of hypoglossal nerve size cannot be inferred simply based on the size of the hypoglossal canal. This is especially pertinent when attempting to draw conclusions from hominin fossil cranial remains, as the canal is the only possible analogue for hypoglossal canal

dimensions. Therefore, reliable conclusions on articulatory control of the tongue cannot be made off of these dimensions without a fair degree of speculation.

The Relationship Between Speech and the Descent of the Larynx

In the past, the descent of the larynx was proposed to be a uniquely human trait directly linked to the development of speech after infancy (the age range from birth to six years of age) (Nishimura, 2006). As humans age, the larynx rapidly descends, which lengthens the VTv and allows the pharyngeal cavity to serve as an audible vocal space. To verify the connection between language production and laryngeal descent, Nishimura et al. performed a study that compared the development of the vocal tracts in three living chimpanzees (Nishimura et al., 2006). Their findings clearly demonstrated that chimpanzees also experience developmental laryngeal descent after birth. This suggests that the adaptation arose before the evolutionary divergence from our last common ancestor with chimpanzees, and likely did not originally have a purpose related to speech. Research by Fitch and Reby (2001) on red deer also demonstrates that the descent of the larynx is not uniquely human. Their research indicates that laryngeal descent primarily serves to lower the resonance frequency of their calls, which in turn may exaggerate the perceived body size to the listener. Nonetheless, while laryngeal descent may serve a function in modern human speech, the adaptation likely arose for a different purpose.

The Relationship Between Speech and Hyoid Morphology

The relationship between the hyoid and speech anatomy is fairly well documented, and comparisons of hyoid structure between chimpanzees, extinct hominins, and modern humans

have been particularly insightful (Steele, 2013). A small cup-like indentation on the hyoid called the hyoid bulla can indicate the presence of air sacs in primates (*Figure 4*). In non-human apes, air sacs amplify the volume of low frequencies, which allow calls to travel further through dense foliage. However, hominin hyoid anatomy suggests that this feature was not positively selected in recent human evolution (Steele, 2013). Acoustic comparisons between simulated vocal tract models with air sacs and those without them reveal that air sacs add an additional frequency to vocal utterances which can obscure the primary frequencies of the vocal tract, and limit the amount of unique, distinguishable sounds that can be made (Boer, 2012). This implies that as humans evolved, the ability to produce clearly distinguishable vocal sounds became more important than producing calls that can travel further distances.

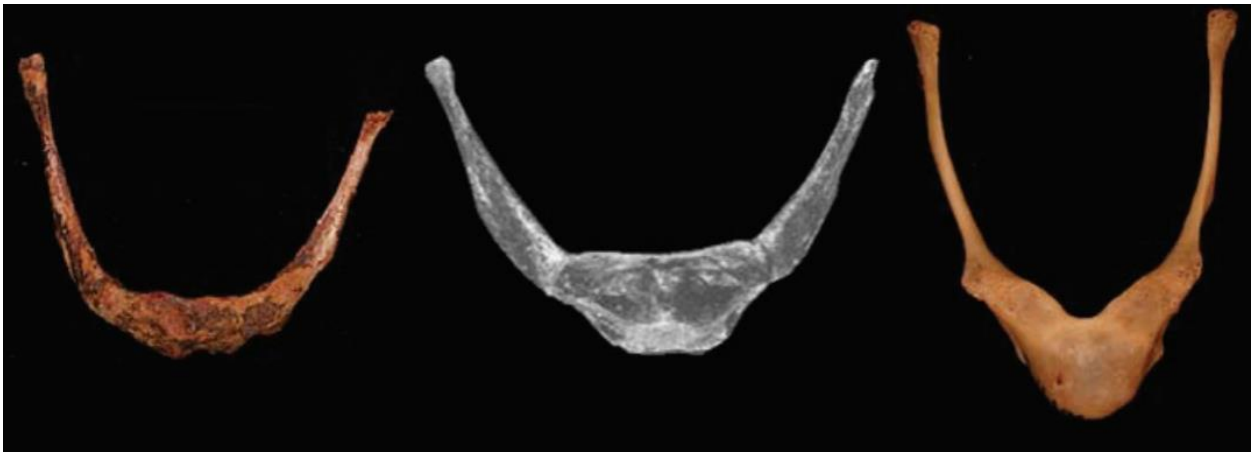


Figure 4. From left to right, Modern H. sapiens, H. neanderthalensis, and Chimpanzee Hyoids.

SOURCE: (Steele, 2013. Figure 1, Page 641)

Air sacs are present in chimpanzees, and the existence of a hyoid bulla in *Australopithecus afarensis* likely correlates to the presence of air sacs as well (Steele, 2013). *H. neanderthalensis* and *H. heidelbergensis* lack a hyoid bulla, so it is probable that the loss of air

sacs occurred at least after the emergence of *H. heidelbergensis* 600,000 years ago (Martinez, 2007). If the disappearance of air sacs is an adaptation primarily to improve the clarity of speech, it is possible that *H. heidelbergensis* or even earlier hominin species were undergoing adaptations with a primarily speech-related purpose. This interpretation is substantiated by analyzing other adaptations related to speech that arose during the same time period.

The Relationship Between Speech and Vertebral Morphology

An article by MacLarnon and Hewitt (1999) details the relationship between the size of the thoracic vertebral canal and breath control during speech. The thoracic section of the spinal cord innervates the muscles that regulate the intercostals and the abdominal muscles, but also the control of breath (lung contraction) while speaking. There is a marked increase in thoracic vertebral canal size in *H. sapiens* and *H. neanderthalensis* when compared with *H. ergaster* and other earlier hominins (MacLarnon, 1999). While explanations for this change have been proposed, such as the emergence of bipedalism, enhanced breathing for endurance running, prevention of choking, and abdominal support during childbirth, none of them adequately account for the adaptation, whether due to inaccurate timing or an insufficient demand for increased innervation (MacLarnon, 1999).

The precise breath control that humans exhibit during speech is quite demanding neurologically however. While speaking, humans are capable of precisely manipulating their lungs, intercostals and abdominal muscles to control the pitch, volume and duration of utterances, while taking sharp inhalations during intentional pauses (MacLarnon, 1999). Unlike the hypoglossal canal, the dimensions of the thoracic vertebral foramen correlate to the size of

the spinal cord, so it is possible to use fossil remains of the thoracic vertebrae as an analogue for the spinal cord itself.

MacLaron and Hewitt (1999) indicate that this adaptation would have emerged no earlier than 1.6 Mya, as *H. ergaster* did not have an enlarged thoracic vertebral canal. At its latest, the adaptation could have emerged alongside the emergence of *H. neanderthalensis* and modern humans around 300,000 years ago, as both species do have increased canal size. It is difficult to more accurately discern the time this adaptation emerged due to the scarcity of thoracic vertebrae for other intermediate hominin species. Unfortunately, the presence of an enlarged thoracic vertebral canal in *H. heidelbergensis* cannot be substantially verified until more complete *H. heidelbergensis* thoracic vertebrae are uncovered.

Available Homo heidelbergensis Fossils

Extensive research on the anatomical traits related to vocalization was not possible for *H. heidelbergensis* until sufficient fossil specimens were discovered. Ninety percent of *H. heidelbergensis* fossil remains were found from 1983 and onwards at the Sima de los Huesos archaeological site in Spain. This site has yielded over 5,500 full or partial fossil specimens attributed to *H. heidelbergensis*, which were dated to at least 350,000 years ago. Most importantly to the study of the evolution of language are two hyoid bones, the first discovered in 1994 and the second discovered in 1997 (Martinez, 2007). Also important to the study of *H. heidelbergensis* vocal anatomy are the four cervical vertebrae that were discovered at Sima de los Huesos (C1, C3, C4, and C7) *(Figure 5)* and the three cervical vertebrae that were discovered

at Gran Dolina (C1, C3, and C7) (Carretero, 1999). Now that these fossils are available to researchers, direct research into the *H. heidelbergensis* vocal anatomy is possible.

Figure 5. Photograph of Sima de los Huesos H. heidelbergensis vertebrae.

(a) Superior and (b) inferior views of the SH atlas AT-1554. (c) Superior and (d) lateral views of the SH cervical 3 AT-1559. (e) Superior and (f) lateral views of the SH cervical 7 SH C7-I. (g) Superior and (h) lateral views of the SH cervical 4 AT-1557. Scale bar represents 5 cm.

SOURCE: (Martinez, 2007. Figure 2, Page 465)

CHAPTER 3: METHODOLOGY

As previously stated, the intent of this research was to reconstruct the vowel space of *Homo heidelbergensis*. This was done by measuring a sample of five *H. heidelbergensis* specimens and inputting the dimensions of the VTh and VTv into the software VTCalcs. VTCalcs is a program that accepts VT length and area dimensions as an input, and outputs the formant frequencies that would be produced with that configuration. The measurements for Atapuerca 5 were acquired through the three dimensional imaging software MeshLab. The measurements for Arago 21, Petralona 1, Kabwe 1, and Steinheim were predicted through multiple linear regression equations. The multiple linear regression equations for this process were created by collecting linear measures from CT images from a sample of 11 modern human individuals.

Five *H. heidelbergensis* crania were measured for this research (Table 1). The Arago 21 and Petralona 1 specimens were acquired from the publicly published three dimensional models by the Smithsonian Institute. The Kabwe 1 cranium was provided by the online database morphosource.com. The University of Central Florida (UCF) provided accurately scaled teaching specimens of the Atapuerca 5 cranium, the Atapuerca 5 mandible, and the Steinheim cranium to use for this research. These specimens were processed into three dimensional images through photogrammetry. The decision to measure the UCF teaching specimens through three dimensional imaging software was made to increase accuracy of the landmarks measured when compared to manually measuring the physical specimens themselves, as well as to maintain consistency in the method of measurement, since Arago 21, Kabwe 1, and Petralona 1 were provided for this research as three dimensional models.

Table 1

Specimens Selected for Measurement			
Name	**Source**	**Estimated Age**	**Location of Origin**
Arago 21 - Cranium	(3D Model) Smithsonian Institution	450,000 ya	Tautavel, France
Atapuerca 5 - Cranium & Mandible	UCF Teaching Specimen	500,000-350,000 ya	Sierra de Atapuerca, Spain
Kabwe 1 - Cranium	(3D Model) morphosource.com	300,000-125,000 ya	Kabwe, Zambia
Petralona 1 - Cranium	(3D Model) Smithsonian Institution	350,000 - 150,000 ya	Petralona, Greece
Steinheim - Cranium	UCF Teaching Specimen	350,000-250,000 ya	Steinheim, Germany

Photogrammetry

Three dimensional models of the UCF teaching specimens were created through photogrammetry due to its low-cost and relative accuracy when compared to other three dimensional imaging methods such as light scanning or laser scanning (Moraes, 2014). Photogrammetry utilizes the information gathered from photographs to create detailed, accurate three dimensional images. Photographs were taken in three to four rings encircling the object, with each ring elevating the camera's position to gather data from higher positions on the object, and with each individual photo capturing at least 85% of the previous photograph's view (Figure 6).

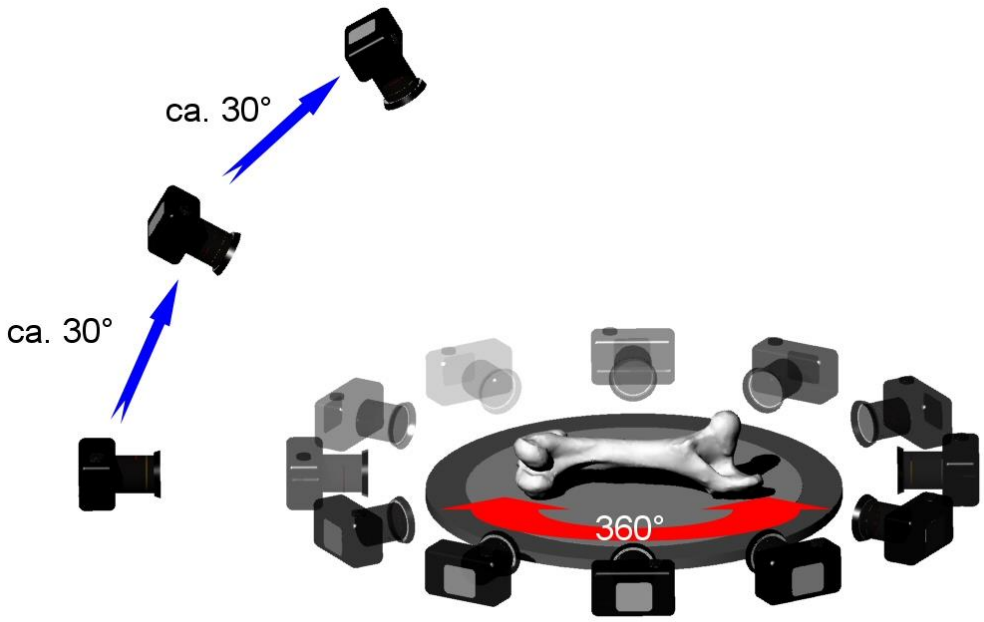


Figure 6. Representation of Camera Positions and Angle Changes for Photogrammetry

SOURCE: (Mallison, 2013)

To capture these images, I used a Canon EOS Rebel 5 with corded remote control, positioned on a tripod. The specimens were placed in a Kodak portable lightbox to ensure that all images were taken with consistent lighting. Rather than circling the camera around the specimen, the specimen was placed on a turntable to gather photographs of the object without needing to move the camera, other than to adjust the height of the tripod for each ring of photos. To capture images of all of the object's surfaces, two sets of photos were taken for each specimen: one set with the superior side facing upwards and one set with the inferior side facing upwards. In total, 590 photographs were taken, with approximately 200 photographs taken for each specimen, divided into two sets of approximately 100 photographs per set.

Three Dimensional Model Processing

To render the three dimensional images, I utilized the software Agisoft PhotoScan Professional 1.3.4. The following procedure was completed independently for each UCF teaching specimen. First, one set of photographs for the specimen, known as a "chunk" in Agisoft, was uploaded to the software. The photographs were then aligned using the Align Photos option with the highest accuracy in order to create a sparse point cloud. From this, a dense point cloud was created using the Build Dense Cloud option with the highest accuracy and mild depth filtering. Creating the dense point cloud with mild depth filtering prompted the program to render the fine details on the specimen, which is useful for preserving anatomical landmarks. Lastly, the model was rendered from the dense point cloud by using the Build Mesh option, and the color and shading was applied by selecting Build Texture. Since each specimen had two sets of photographs, the second set of photographs was rendered by selecting Add Chunk, and repeating the aforementioned steps for that chunk.

The product of this process was two 3D models, one with the superior side facing upwards (and the inferior side obscured) and one with the inferior side facing upwards (and the superior side obscured) (See Figures 7 and 8 for images of these two models). To merge the two models into one measurable object, each model was first cropped by using the select tool and deleting the obscured portion of the model (See Figures 9 and 10 for images of the two cropped models). Then, the two chunks were aligned using the Align Chunks option with the highest accuracy and the Point Based method of alignment. This alignment method aligns both models by directly analysing the point clouds of each chunk, which produced a more accurate result than Landmark Based alignment. Lastly, the two models were merged into one model using the

Merge Chunks option and selecting Merge Dense Clouds and Merge Models. This process was done for the Atapuerca 5 cranium, the Atapuerca 5 mandible, and the Steinheim cranium. The product of this process was a single three dimensional model for each specimen (See Figures 11, 12, and 13 for images of the final Atapuerca 5 model).

Figure 7. Atapuerca 5 Superior Side Up - Initial 3D Model

Figure 8. Atapuerca 5 Inferior Side Up - Initial 3D Model

Figure 9. Atapuerca 5 Inferior Section Cropped

Figure 10. Atapuerca 5 Superior Section Cropped

Figure 11. Atapuerca 5 Final 3D Model - Anterior

Figure 12. Atapuerca 5 Final 3D Model - Diagonal

Figure 13. Atapuerca 5 Final 3D Model - Lateral

Model Scaling and Measurements

To measure the three dimensional models, I utilized the software MeshLab 2016. When a model is rendered in Agisoft PhotoScan, the scale of the model is set to an arbitrary value which does not necessarily reflect the object's actual size. Thus before measurements were made, I utilized MeshLab's scaling feature "Transform: Scale, Normalize" to correct the object's scale. This allows users to input the ratio of a known linear distance divided by MeshLab's arbitrary value for that distance. For instance, the initial linear measurement for the orbit height of the Atapuerca 5 cranium was an arbitrary value of 1.26. By dividing the actual orbit height in millimeters (33mm) by this arbitrary value of 1.26, I received the scaling factor that I inputted into MeshLab's scaling feature to correct the object's scale.

To maintain consistency when scaling each specimen, I used the orbit height and orbit breadth from Rightmire's measurements of Middle Pleistocene *Homo* crania for each specimen (2008). This process involved measuring the orbit height to receive MeshLab's arbitrary value, dividing the actual orbit height in millimeters from Rightmire's measurements by MeshLab's arbitrary value, scaling the model with Transform: Scale, Normalize by inputting the scaling factor from that ratio, then checking for accuracy by measuring orbit height again, and subsequently measuring the orbit breadth. The orbit height and orbit breadth were used to scale these images because they were consistent, easily locatable linear distances that were present on all specimens measured.

I used the same process to scale the Atapuerca 5 mandible by utilizing Rosas' measurements of the Atapuerca 5 mandible (1997). The linear distance between the anatomical landmark gonion and the apex of the coronoid process was the primary measure used to scale the

mandible. Accuracy was checked by measuring the linear distances between that primary measure, as well as two additional measures: the linear distance from gnathion to the apex of the mandibular condyle and the distance from gonion to the apex of the mandibular condyle.

To estimate the vowel space of *Homo heidelbergensis*, the key linear distances that must be known are the lengths of the oral cavity (VTh) and the pharyngeal cavity (VTv). A study on vocal tract reconstruction by Boë et al. established that these two distances can be estimated by measuring the distances between skeletal landmarks (2013). The length of the VTh can be estimated by measuring the distance between prosthion and the anterior tubercle of C1 (specifically, the most anterior point on the midline of the anterior tubercle of C1). These skeletal landmarks represent the distance between lips and the area of contact between the velum and pharyngeal wall. The VTv can be estimated by measuring the distance between the anterior tubercle of C1 and the most inferior point on the anterior surface on the midline of C5. These skeletal landmarks represent the distance between the area of contact between the velum and pharyngeal wall, and the center of the range of variation for the average vertical position of the glottis. Boë et al. demonstrated that the normal vertical position of the modern human glottis falls between the superior limit of C5 and the inferior limit of C6 (2013). Thus, I defined the most inferior point on the anterior surface on the midline of C5 as the approximate vertical position of the glottis for all specimens, as this skeletal landmark represents the approximate center of the range of the expected glottis vertical position.

To measure these distances on the Atapuerca 5 specimen, a three dimensional model of a spine from a modern human male cadaver was used as a proxy for the *Homo heidelbergensis* spine, which was scaled to match *Homo heidelbergensis* spinal measurements from Carretero et

al. (1999). This spine model was obtained from the database of publicly available 3D models, Grabcad.com. An et al. obtained the initial CT slices for the model through radiography to ensure that the integrity of the vertebra was normal and that there were no pathological abnormalities (2014). An et al. then converted these CT slices into the solid model that I obtained for this research.

In previous research on the *H. neanderthalensis* vowel space, Barney et al. obtained measurements from a modern human spine to serve as a proxy for the *H. neanderthalensis* spine (2012). This was deemed to provide a reliable approximation of the *H. neanderthalensis* VTv length, as all known cervical vertebrae from the Middle Pleistocene onwards have been measured to be within the modern human range of variation (Carretero, 1999). Nonetheless, to ensure that length of the *Homo heidelbergensis* VTv was measured properly for this research, the three dimensional spine model from An et al. was scaled to match the measurements of the cervical spinal vertebrae discovered in Sierra de Atapuerca (Carretero, 1999). In this study by Carretero et al., the dimensions of the cervical vertebrae from Atapuerca were shown to be within the modern human range. I used the same process to scale the spine as I did with the Atapuerca 5 cranium and mandible. The maximum transverse diameter of C1 from Carretero et al. was the primary value used to obtain the scaling factor. Accuracy was checked by measuring the maximum dorsoventral diameter of C1, the maximum height of the lateral masses of C1, the maximum superior transverse diameter of C3, and the maximum superior transverse diameter of C4. After scaling the model, each of these measurements were found to be within 0.5mm of the measurements by Carretero et al..

I imported the scaled three dimensional models of the Atapuerca 5 cranium, the Atapuerca 5 mandible and the modern human spine into one workspace. I used Meshlab's Translate and Rotate options to orient the models into their anatomically correct positions. The jaw was oriented in a closed position. The method of vowel space reconstruction used in this research did not require the use of a hyoid to estimate the dimensions of the VT. My process for estimating VT area without the hyoid is provided below under the *Formant Frequency Analysis* subheading. Nonetheless, a three dimensional model of a modern human hyoid from the public database 3dprint.nih.gov was scaled and oriented with the Atapuerca 5 skull to serve as a visual reference (Leggett, 2014). I used measurements of the *Homo heidelbergensis* hyoid AT-1500 from Sierra de Atapuerca to scale the modern human hyoid (Martinez, 2007). The maximum transverse diameter of the hyoid body was the primary value used to obtain the scaling factor. Accuracy was checked by measuring the maximum transverse diameter of the hyoid body and the maximum medial height of the hyoid body. The position of the hyoid was estimated by aligning the hyoid with the intersection of three planes: the transverse plane drawn from the mentum, the coronal plane drawn from the mandibular foramen, and the midline. This method was established to be an accurate method of estimating hyoid position by Granat et al. (2007).

After scaling and orientation were complete, the Atapuerca 5 cranium and mandible, the modern human spine, and the modern human hyoid were merged into a single three dimensional model in MeshLab. See Figures 14, 15, and 16 for images of this model. Note that a texture was not applied. As previously discussed, Boë et al. established that the VTh can be approximated by measuring the linear distance of prosthion to the anterior tubercle of C1, and the VTv can be approximated by measuring the linear distance of the anterior tubercle of C1 to the most inferior

point on the anterior surface of C5. These two linear distances were measured in MeshLab by utilizing MeshLab's measure tool, which requires the user to identify two points that act as endpoints for a linear distance measurement. The measurements were taken on two separate occasions with at least 24 hours in between measurement trials to avoid memory bias of measured distances. Results of these measurements are in the Results section.

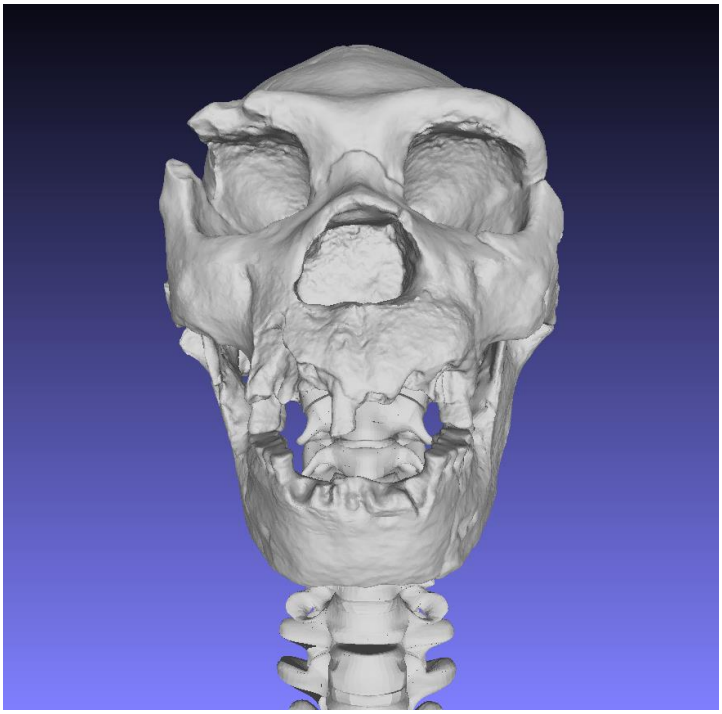


Figure 14. Atapuerca 5 Cranium, Mandible, Spine, and Hyoid - Anterior

Figure 15. Atapuerca 5 Cranium, Mandible, Spine, and Hyoid - Diagonal

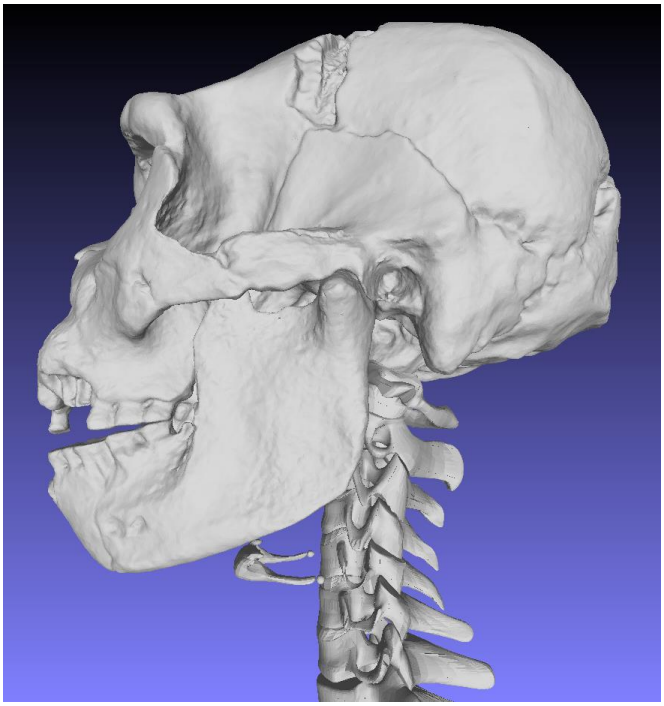


Figure 16. Atapuerca 5 Cranium, Mandible, Spine, and Hyoid - Lateral

Computed Tomography Scan Measurements

The Atapuerca 5 cranium and mandible was measured by using a spine proxy: a modern human spine that was scaled to match the dimensions of *Homo heidelbergensis* vertebrae from Sierra de Atapuerca. However, four additional craniums were measured for this research, Arago 21, Kabwe 1, Petralona 1, and the Steinheim cranium. These craniums originate from various locations across Europe and Africa from 500,000 - 125,000 years ago. No *Homo heidelbergensis* cervical vertebrae have currently been found outside of Spain, so rather than using the same spine proxy for each specimen, a different method was used to obtain the vocal tract dimensions which better accounted for variation in fossil age and location of origin. To predict the VTh and VTv lengths for these specimens, measures on the cervical spine are required. To predict these measures that I could not take directly on the craniums, I built regression equations with the linear distances between 12 craniospinal landmarks on a sample of 11 modern human individuals. To measure these individuals, CT scans were provided by Dr. William Sensakovic, Department of Radiology, Florida Hospital. Measurements were taken in the program Amira 6.4.

The following 12 landmarks were used for each individual. Landmarks were placed directly on an orthoslice set on the midline of each individual (Figure 17).

Skull

1. INF: Infradentale
2. PRO: Prosthion
3. PMP: Midpoint on the Inferior Surface of the Palate
4. ANS: Anterior Nasal Spine
5. PNS: Posterior Nasal Spine

Spine

6. C1T: Most Anterior Point of the Anterior Tubercle of Cervical 1
7. C4S: Most Superior Point on the Anterior Surface of Cervical 4
8. C4I: Most Inferior Point on the Anterior Surface of Cervical 4
9. C5S: Most Superior Point on the Anterior Surface of Cervical 5
10. C5M: Midpoint on the Anterior Surface of Cervical 5
11. C5I: Most Inferior Point on the Anterior Surface of Cervical 5
12. C6S: Most Superior Point on the Anterior Surface of Cervical 6

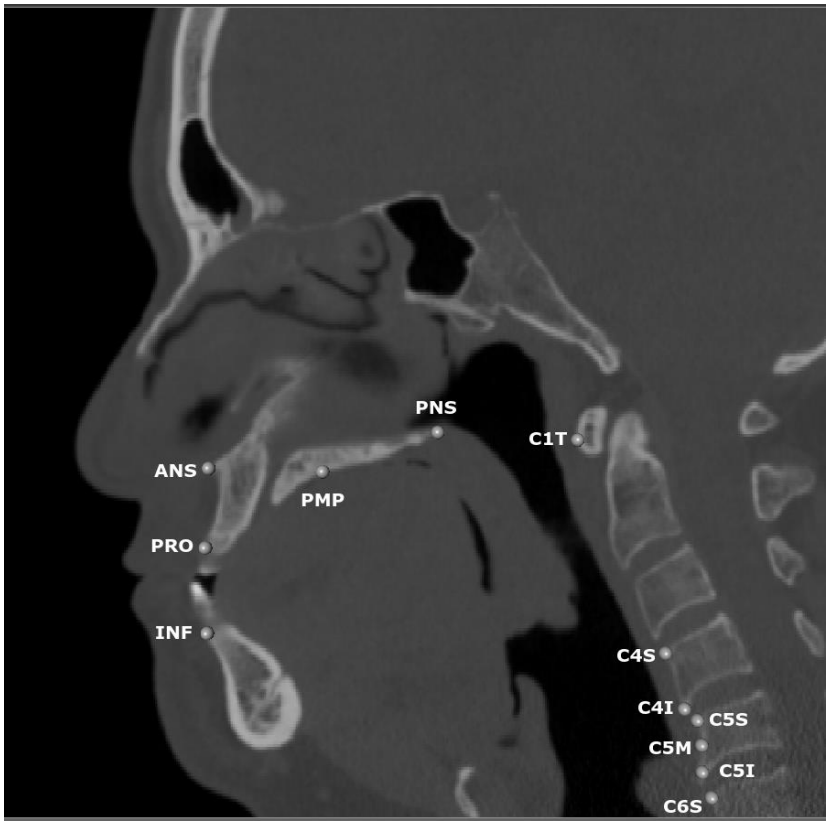


Figure 17. Landmarked CT Image Orthoslice

These landmarks were selected because they are easily identifiable points that represent the outer boundaries of the skeletal portions of the oral cavity, nasal cavity, and pharyngeal cavity.

Regression Equations

Anatomical landmark coordinates were converted into XYZ files and input into WinEDMA software (v 1.0.1). Using the utility and convert options, these files were converted into a file consisting of all unique inter-landmark linear distances for the measured structures. EDMA software uses a standard distance formula derived from the Pythagorean theorem that takes as input either x,y or x,y,z coordinate values. The anatomical coordinates for this project are in three-dimensions (3D). Consequently, the linear distance between any two anatomical landmark coordinates with coordinate values of, for example, A = x_1,y_1,z_1 and B = x_2,y_2,z_2 can be calculated using the following formula:

$$d_{(A,B)} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (x_1 - x_2)^2}$$

The number of unique inter-landmark linear distance measurements that can be calculated from a set of landmark coordinate values is:

$$\frac{n\,(n-1)}{2},$$

where n is equal to the number of landmark coordinates input into the program. For further EDMA details, readers are invited to review Lele and Richtsmeier (2001). WinEDMA labels calculated linear distances by listing the two anatomical landmark endpoints and connecting

them by an "&" symbol. A total of 66 linear distances were calculated from the 12 anatomical landmark coordinates that were measured on each of the 11 individuals (See Table 2).

Table 2

Linear Distances								
PRO&INF	PMP&INF	**PMP&PRO***	ANS&INF	**ANS&PRO***	**ANS&PMP***	**PNS&PRO***	**PNS&PMP***	**PNS&ANS***
18.109	32.991	**25.406**	30.004	**12.903**	**20.46**	**53.371**	**28.093**	**47.652**
18.527	35.683	**24.225**	40.329	**22.151**	**20.619**	**55.643**	**33.082**	**51.842**
28.971	45.738	**28.106**	48.943	**20.528**	**23.807**	**52.595**	**25.374**	**48.038**
30.685	41.486	**23.487**	43.358	**12.867**	**22.388**	**55.855**	**33.361**	**55.516**
20.551	36.895	**27.504**	40.022	**20.035**	**25.436**	**54.562**	**28.53**	**53.28**
21.65	43.795	**29.175**	39.479	**18.047**	**22.262**	**56.42**	**28.138**	**49.794**
20.243	43.505	**29.805**	43.545	**23.486**	**23.22**	**57.056**	**30.449**	**53.318**
20.463	42.902	**32.648**	41.414	**21.788**	**25.666**	**56.747**	**24.878**	**49.555**
21.061	34.919	**20.365**	40.866	**19.931**	**24.283**	**48.599**	**29.17**	**51.712**
19.446	41.595	**28.38**	36.031	**16.657**	**22.675**	**55.753**	**27.716**	**49.368**
14.799	32.722	**26.222**	33.065	**19.405**	**19.53**	**53.313**	**29.708**	**48.609**
PNS&INF	C1T&INF	**C1T&PRO****	C1T&PMP	C1T&ANS	C1T&PNS	C4S&INF	C4S&PRO	C4S&PMP
57.256	87.808	**88.329**	64.083	84.314	37.17	85.2	95.193	79.422
60.62	81.885	**84.458**	66.677	87.087	38.004	111.714	121.83	111.687

62.895	88.494	**85.153**	59.41	82.466	34.435	77.18	92.339	80.129
61.966	86.618	**87.666**	66.306	88.654	33.404	84.63	101.485	88.243
57.01	82.968	**84.01**	58.935	83.653	30.511	83.218	95.763	81.978
66.477	96.348	**91.602**	65.504	87.663	38.346	94.735	101.121	85.987
63.942	86.68	**85.739**	63.11	86.32	33.774	92.522	102.996	94.677
62.696	94.554	**94.321**	64.91	90.276	40.955	114.601	123.291	102.814
56.692	74.078	**71.686**	54.269	77.921	26.838	96.733	106.416	96.982
65.942	90.905	**84.317**	57.568	79.878	30.837	99.819	103.203	85.551
54.005	81.742	**84.742**	63.637	82.642	34.428	88.337	97.917	86.506
C4S&ANS	C4S&PNS	C4S&C1T	C4I&INF	C4I&PRO	C4I&PMP	C4I&ANS	C4I&PNS	C4I&C1T
97.929	65.294	48.924	88.049	99.621	85.945	103.535	74.182	59.543
131.822	90.255	55.414	118.692	129.766	120.825	140.742	100.499	66.12
101.863	67.584	56.104	81.047	99.642	90.501	111.114	80.158	70.008
108.322	66.25	48.981	89.369	107.995	95.813	115.389	74.856	57.522
105.759	63.59	51.99	88.954	103.162	91.521	114.553	74.749	63.792
106.074	69.666	49.523	94.669	103.151	90.618	109.699	76.746	59.356
115.106	76.155	54.644	97.326	109.767	104.34	123.73	87.907	67.779
127.666	84.894	53.305	121.125	131.238	112.566	136.977	95.754	65.296
120.557	79.065	57.865	107.044	117.56	108.617	132.075	90.816	69.113

106.827	69.219	50.504	104.848	110.129	94.588	115.133	80.12	62.539
104.415	65.495	44.695	93.735	104.411	95.104	112.426	75.814	56.143
C4I&C4S	C5S&INF	C5S&PRO	C5S&PMP	C5S&ANS	C5S&PNS	C5S&C1T	C5S&C4S	C5S&C4I
10.633	92.086	104.122	90.99	108.349	79.546	64.347	15.546	5.383
10.783	122.864	134.006	125.076	144.993	104.56	69.911	14.507	4.297
13.917	82.232	102.065	94.186	114.281	84.793	75.45	19.346	5.493
8.819	90.088	110.128	99.018	118.031	79.46	63.335	14.375	6.045
11.876	95.002	109.568	98.164	121.15	81.209	69.197	17.925	6.744
10.188	97.773	107.424	96.224	114.774	83.327	66.332	16.928	7.034
13.192	102.468	115.369	110.453	129.683	94.098	73.453	18.835	6.212
12.026	126.335	136.718	118.258	142.635	101.425	70.457	17.198	5.704
11.793	110.896	121.472	112.49	135.971	94.462	72.421	15.574	3.912
12.053	107.441	113.139	98.017	118.416	83.807	66.167	15.663	3.718
11.452	95.025	106.005	97.353	114.468	78.63	59.443	14.749	3.333
C5M&INF	C5M&PRO	C5M&PMP	C5M&ANS	C5M&PNS	C5M&C1T	C5M&C4S	C5M&C4I	C5M&C5S
93.732	106.483	94.449	111.279	84.192	69.995	21.11	10.581	5.723
125.66	137.371	129.243	148.992	109.585	75.411	20.059	9.293	5.992
87.262	108.07	100.99	120.781	91.904	82.12	26.101	12.219	7.118

93.454	114.662	104.389	122.955	85.704	69.874	20.933	12.41	6.563
99.003	113.964	103.031	125.852	86.284	74.03	23.016	11.748	5.109
100.858	110.972	100.217	118.602	87.493	70.107	20.591	11.059	4.167
105.68	119.243	115.347	134.18	99.674	79.398	24.767	11.861	5.95
131.12	142.008	124.167	148.385	107.64	76.792	23.558	11.907	6.362
115.895	126.706	117.836	141.294	99.753	77.481	20.929	9.222	5.356
110.091	116.56	102.341	122.393	88.887	71.717	21.216	9.184	5.574
100.071	111.224	102.757	119.857	83.931	63.917	19.398	8.19	5.47
C5I&INF	C5I&PRO	C5I&PMP	C5I&ANS	C5I&PNS	**C5I&C1T****	C5I&C4S	C5I&C4I	C5I&C5S
95.313	108.606	97.45	113.844	88.111	**74.66**	25.749	15.153	10.442
128.377	140.296	132.439	152.129	112.995	**78.878**	23.519	12.76	9.292
89.078	110.901	104.959	124.244	96.704	**87.637**	31.564	17.647	12.301
94.727	117.128	107.872	125.859	90.468	**75.68**	26.699	18.381	12.393
102.895	118.302	107.976	130.567	91.594	**79.345**	28.407	16.993	10.484
101.346	112.296	102.669	120.556	90.943	**74.629**	25.175	15.322	8.302
110.076	124.246	121.274	139.742	106.151	**86.046**	31.424	18.487	12.593
134.588	146.167	129.327	153.252	113.432	**83.35**	30.049	18.13	12.932
117.71	129.417	121.409	144.578	104.401	**82.851**	25.352	13.746	10.59
111.288	118.741	105.816	125.327	93.6	**77.548**	27.171	15.158	11.795

102.72	114.412	107.227	123.891	89.522	**70.54**	25.882	14.445	11.231
C5I&C5M	C6S&INF	C6S&PRO	C6S&PMP	C6S&ANS	C6S&PNS	C6S&C1T	C6S&C4S	C6S&C4I
4.72	98.709	112.25	101.421	117.666	92.221	78.382	29.538	19.022
3.467	130.692	142.936	135.554	155.133	116.594	82.731	27.45	16.667
5.594	93.005	115.473	110.156	129.177	102.193	93.002	36.982	23.081
6.068	97.176	120.196	111.438	129.139	94.532	79.905	30.924	22.542
5.399	107.513	123.051	112.761	135.364	96.185	83.295	32.905	21.715
4.785	101.545	113.464	105.214	122.466	94.727	79.808	30.528	20.468
6.66	112.866	127.387	124.973	143.214	110.187	90.206	35.586	22.631
6.773	137.993	149.607	132.715	156.672	116.686	86.168	32.928	21.154
6.563	121.723	133.463	125.405	148.602	108.181	86.34	29.119	17.383
6.412	115.193	123.16	110.797	130.084	98.916	82.839	32.397	20.348
6.823	106.771	118.459	111.095	127.858	93.017	73.149	28.722	17.472
C6S&C5S	C6S&C5M	C6S&C5I						
14.038	8.443	4.112						
13.334	7.421	4.049						
17.797	10.886	5.523						
16.592	10.147	4.236						

15.076	9.979	4.859						
13.61	10.454	5.687						
16.754	10.821	4.162						
15.73	9.377	3.441						
13.919	9.053	4.047						
16.842	11.285	5.365						
14.475	9.328	4.094						

Values in mm.

***Linear Distance Used to Carry Out a Multiple Linear Regression**

****Linear Distances to be Predicted by the Multiple Linear Regression**

Far more linear distance values were generated than were needed for this investigation. The following linear distance measurements were inputted into SPSS (v24) to carry out a multiple linear regression: PMP&PRO, ANS&PRO, ANS&PMP, PNS&PRO, PNS&PMP, and PNS&ANS. These linear distances were selected because they consist of cranial landmarks from the set of 12 landmarks taken on the *Homo sapiens* CT images which were able to be measured on the *Homo heidelbergensis* specimens. These linear distances were used to generate prediction equations for the following linear dimensions needed to estimate vowel space: C1T&PRO, and C5I&C1T.

For the first multiple linear regression, the C1T&PRO linear distance measurement was input as the dependent variable, while PMP&PRO, ANS&PRO, ANS&PMP, PNS&PRO,

PNS&PMP, and PNS&ANS linear distance measures were input as the independent variables. See Table 3 for correlations between dependent and independent variables.

Table 3

Correlations							
		PMP_PRO	ANS_PRO	ANS_PMP	PNS_PRO	PNS_PMP	PNS_ANS
Pearson Correlation	C1T_PRO	.748	-.131	-.046	.797	-.249	-.197
Sig. (1-tailed)	C1T_PRO	.004	.350	.447	.002	.230	.281

The results of the multiple linear regression suggest that a non-significant proportion of the total variation in C1T&PRO was predicted by the independent variables (F (6, 4) = 4.233, p-value = 0.092). See Table 4 for Analysis of Variance.

Table 4

ANOVA						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	279.754	6	46.626	4.233	.092b
	Residual	44.058	4	11.015		
	Total	323.812	10			
a. Dependent Variable: C1T_PRO						
b. Predictors: (Constant), PNS_ANS, ANS_PRO, PNS_PRO, ANS_PMP, PMP_PRO, PNS_PMP						

The intercept (or average PRO&INF value when all independent variables are zero) was 48.776 and not statistically significantly different from zero (t = 1.068, df = 4, p-value = 0.346). None of the independent variables were significantly different from zero. Their unstandardized coefficients (test statistic, p-values) are as follows: PMP&PRO = -6.034 (t = -0.712, p-value =

0.516), ANS&PRO = 1.359 (t = 0.577, p-value = 0.595), ANS&PMP = -10.035 (t = -0.779, p-value = 0.480), PNS&PRO = 7.259 (t = 0.969, p-value = 0.388), PNS&PMP = -14.747 (t = -0.786, p-value = 0.476), and PNS&ANS = 8.518 (t = 0.737, p-value = 0.502). See Table 5 for coefficients.

Table 5

Coefficients							
Model		Unstandardized Coefficients		t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error			Lower Bound	Upper Bound
1	(Constant)	48.776	45.690	1.068	.346	-78.080	175.632
	PMP_PRO	-6.034	8.473	-.712	.516	-29.559	17.491
	ANS_PRO	1.359	2.355	.577	.595	-5.179	7.897
	ANS_PMP	-10.035	12.882	-.779	.480	-45.802	25.733
	PNS_PRO	7.259	7.493	.969	.388	-13.546	28.064
	PNS_PMP	-14.747	18.761	-.786	.476	-66.835	37.341
	PNS_ANS	8.518	11.557	.737	.502	-23.570	40.607
a. Dependent Variable: C1T_PRO							

Multiple R^2 indicates that 86.4% of the variation C1T&PRO was predicted by the 6 independent variables. From this data, a sample prediction model was generated using the format $Y_i' = b_{X1} X_{1i} + b_{X2} X_{2i} + b_{X3} X_{3i} + b_{X4} X_{4i} + b_{X5} X_{5i} + b_{X6} X_{6i} + a$. The sample prediction model is:

C1T&PRO = -6.034 (PMP&PRO) + 1.359 (ANS&PRO) - 10.035(ANS&PMP) + 7.259(PNS&PRO) -14.747 (PNS&PMP) + 8.518 (PNS&ANS) + 48.776

For the second multiple linear regression, the C5I&C1T linear distance measurement was input as the dependent variable, while PMP&PRO, ANS&PRO, ANS&PMP, PNS&PRO, PNS&PMP, and PNS&ANS linear distance measures were input as the independent variables. See Table 6 for correlations between dependent and independent variables.

Table 6

Correlations							
		PMP_PRO	ANS_PRO	ANS_PMP	PNS_PRO	PNS_PMP	PNS_ANS
Pearson Correlation	C5I_C1T	.239	.601	.688	-.087	-.351	.118
Sig. (1-tailed)	C5I_C1T	.240	.025	.010	.400	.145	.365

The results of the multiple linear regression suggest that a non-significant proportion of the total variation in C5I&C1T was predicted by the independent variables ($F(6, 4) = 1.18$, p-value = 0.457). See Table 7 for Analysis of Variance.

Table 7

ANOVA						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	179.129	6	29.855	1.180	.457b
	Residual	101.220	4	25.305		
	Total	280.348	10			
a. Dependent Variable: C5I_C1T						
b. Predictors: (Constant), PNS_ANS, ANS_PRO, PNS_PRO, ANS_PMP, PMP_PRO, PNS_PMP						

The intercept (or average PRO&INF value when all independent variables are zero) was 38.951 and not statistically significantly different from zero (t = 0.562, df = 4, p-value = 0.604). None of the independent variables were significantly different from zero. Their unstandardized coefficients (test statistic, p-values) are as follows: PMP&PRO = 0.498 (t = 0.039, p-value = 0.971), ANS&PRO = 0.425 (t = 0.119, p-value = 0.911), ANS&PMP = 3.349 (t = 0.171, p-value = 0.872), PNS&PRO = -0.541 (t = -0.048, p-value = 0.964), PNS&PMP = 2.226 (t = 0.078, p-value = 0.941), and PNS&ANS = -1.818 (t = -0.104, p-value = 0.922). See Table 8 for coefficients.

Table 8

Coefficients							
Model		Unstandardized Coefficients		t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error			Lower Bound	Upper Bound
1	(Constant)	38.951	69.254	.562	.604	-153.328	231.230

	PMP_PRO	.498	12.843	.039	.971	-35.159	36.156
	ANS_PRO	.425	3.569	.119	.911	-9.485	10.335
	ANS_PMP	3.349	19.526	.171	.872	-50.865	57.562
	PNS_PRO	-.541	11.358	-.048	.964	-32.076	30.994
	PNS_PMP	2.226	28.436	.078	.941	-76.724	81.177
	PNS_ANS	-1.818	17.518	-.104	.922	-50.455	46.819
a. Dependent Variable: C5I_C1T							

Multiple R^2 indicates that 63.9% of the variation C5I&C1T was predicted by the 6 independent variables. From this data, a sample prediction model was generated using the format $Y_i' = b_{X1} X_{1i} + b_{X2} X_{2i} + b_{X3} X_{3i} + b_{X4} X_{4i} + b_{X5} X_{5i} + b_{X6} X_{6i} + a$. The sample prediction model is:

C5I&C1T = 0.498 (PMP&PRO) + 0.425 (ANS&PRO) + 3.349 (ANS&PMP) - 0.541(PNS&PRO) + 2.226 (PNS&PMP) - 1.818 (PNS&ANS) + 38.951

The following linear distances were measured in MeshLab on the three dimensional models of the Arago 21, Kabwe 1, Petralona 1, and Steinheim craniums: PRO&ANS, PRO&PMP, PMP&PNS, ANS&PNS, PRO&PNS, ANS&PMP. See Table 9 for measurement results.

Table 9

Cranial Measurement Results						
First Measurement Trial						
	PRO&ANS	PRO&PMP	PMP&PNS	ANS&PNS	PRO&PNS	ANS&PMP
Arago 21*	18.09	29.62	24.51	43.36	54.13	15.82
Kawbe 1	35.99	37.32	35.38	54.39	70.65	21.64
Petralona 1	28.64	36.59	36.99	62.58	75.69	27.00
Steinheim	24.79	25.94	36.18	58.12	68.67	25.49
Second Measurement Trial						
Arago 21*	19.02	29.18	25.02	42.89	54.78	16.32
Kawbe 1	35.12	38.16	34.26	54.69	70.31	21.92
Petralona 1	28.32	36.78	36.35	62.88	75.12	27.06
Steinheim	25.12	25.56	36.52	57.96	68.54	25.67
Average of Measurement Trials						
Arago 21*	18.55	29.40	24.76	43.12	54.45	16.07
Kawbe 1	35.55	37.74	34.82	54.54	70.48	21.78
Petralona 1	28.48	36.68	36.67	62.73	75.40	27.03
Steinheim	24.95	25.75	36.35	58.04	68.60	25.58

Values in mm.

***PNS Not Fully Present on Specimen**

The average values of these linear distances were inputted into the two regression equations to predict the linear distances needed to estimate the vowel space: C1T&PRO, and C5I&C1T. See Table 10 for predicted values.

Table 10

Multiple Linear Regression Predicted Values		
	C1T&PRO	C1T&C5I
Arago 21	132.75	62.56
Kawbe 1	113.50	86.02
Petralona 1	135.80	86.64
Steinheim	126.95	86.33

Values in mm.

Linear distance measurements for the Atapuerca 5 specimen were obtained by directly measuring the Atapuerca 5 cranium and scaled modern human spine proxy in MeshLab. See Table 11 for the results of these measurements.

Table 11

Atapuerca Measurements		
Atapuerca	C1T&PRO	C1T&C5I
Trial 1	120.11	81.56
Trial 2	119.82	81.39
Average	119.97	81.48

Values in mm.

Formant Frequency Analysis

The linear distances I obtained through regression equation predictions and direct measurements in MeshLab were used as the lengths of the VTh and VTv for each specimen. As established by Boë et al., the linear distance between C1T&PRO can be interpreted as the approximate length of the VTh, while the linear distance between C1T&C5I can be interpreted as the approximate length of the VTv (2013). These lengths were inputted into the software VTCalcs, a vocal tract model written by Dr. Shinji Maeda to determine the formant frequencies of a given vocal tract configuration. VTCalcs allows users to input the length and area values for the VTh and VTv to receive the formant frequencies, measured in hertz, as an output. The formant frequencies of the three corners of the vowel space, /a/, /i/, and /u/ were the products of this process.

In previous research on vowel space reconstructions, Lieberman et al. established that /a/, /i/, and /u/ are produced by simple configurations of the vocal tract (1972). These vowels are produced by dividing the vocal tract into open and constricted segments, with constricted

segments being approximately 10% of the area of open segments. /a/ is produced with an open VTh and a constricted VTv. /i/ is produced with a constricted VTh and an open VTv. /u/ is produced by using the tongue to divide the VTh and VTv into two halves each. For the oral cavity, the segment of the VTh which is furthest from the glottis (VTh_1) is constricted, and the segment of the VTh which is closest to the glottis (VTh_2) is open. For the pharyngeal cavity, the segment of the VTv which is furthest from the glottis (VTv_1) is constricted, and the segment of the VTv which is closest to the glottis (VTv_2) is open. This creates a vocal tract configuration with four segments, two which are open and two which are constricted. See Figure 18 for a visualization of these configurations.

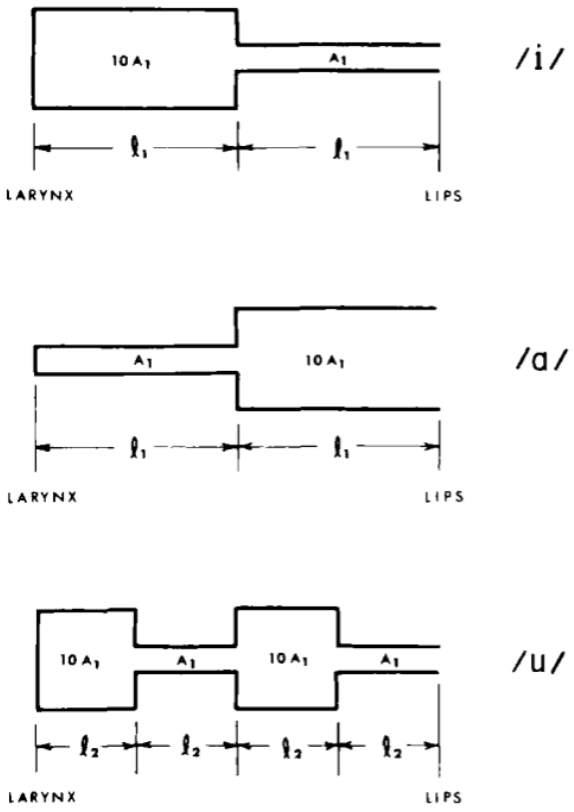


Figure 18. Representation of the Vocal Tract Configurations for /i/, /a/, and /u/.

SOURCE: (Lieberman, 1972)

Because each vowel can be reproduced with these configurations of open and constricted segments, I established a constant area value for open segments and constricted segments based on vocal tract measurements by Story et al. (1996). This study determined that open segments of a vocal tract tend to have an area of approximately 4-6cm^2, and constricted segments of a vocal tract tend to have an area of approximately 0.3-0.7cm^2. I set the average values of these configurations as the constant area values for all area inputs in VTCalcs: 5cm^2 for open segments and 0.5cm^2 for constricted segments.

To obtain the formant frequencies for /a/ and /i/, I used the Two Tube function in VTCalcs. This function allows users to directly input the length and area of the VTh and VTv independently. To obtain the formant frequencies for /u/, I used the Area Function option in VTCalcs. This method was needed due to the vocal tract being separated into four segments for the vowel /u/. The Area Function provides users with the ability to set the total length of the VT, and then divide that VT into any number of sections. The area for each section can then be set individually. As such, I divided the VT into four primary sections: VTh_1, VTh_2, VTv_1, and VTv_2. The area for each section was inputted independently, and the length ratio between VTv:VTh was set for each specimen. See Table 12 for all of the values that were used to obtain the formant frequency outputs.

Table 12

VT Length and Area Values in mm				
	VTh Length	**VTv Length**	**Total VT Length**	**VTv:VTh**
AVG *sapiens*	85.64	79.2	164.84	0.92
Arago*	132.75	62.56	195.31	0.47
Atapuerca	119.97	81.48	201.45	0.68
Kabwe	113.51	86.02	199.53	0.75
Petralona	135.8	86.64	222.44	0.64
Steinheim	126.95	86.33	213.28	0.68
AVG *heidelbergensis*	125.8	80.6	206.4	0.64
AVG *heidelbergensis* ; No Arago	124.06	85.12	209.18	0.68
	VTh Area	**VTv Area**		
/a/	5.0 cm^2	0.5 cm^2		
/i/	0.5 cm^2	5.0 cm^2		
	VTh_1 Area	**VTh_2 Area**	**VTv_1 Area**	**VTv_2 Area**
/u/	5.0 cm^2	0.5 cm^2	5.0 cm^2	0.5 cm^2

***PNS not fully present on specimen**

In total, eight trials were conducted for each vowel, one for each of the five *H. heidelbergensis* specimens, one for the average VT dimensions of all five specimens, one for the average VT dimensions of the *H. heidelbergensis* specimens excluding Arago 21, and one for the average VT dimensions from the modern human sample. The decision to exclude Arago 21 in one of the trials was made to assess output error due to the damaged posterior nasal spine of the specimen. The final formant frequency results did not greatly differ whether Arago 21 was included in the average measurements. See the study limitations section for more information on this decision.

CHAPTER 4: RESULTS

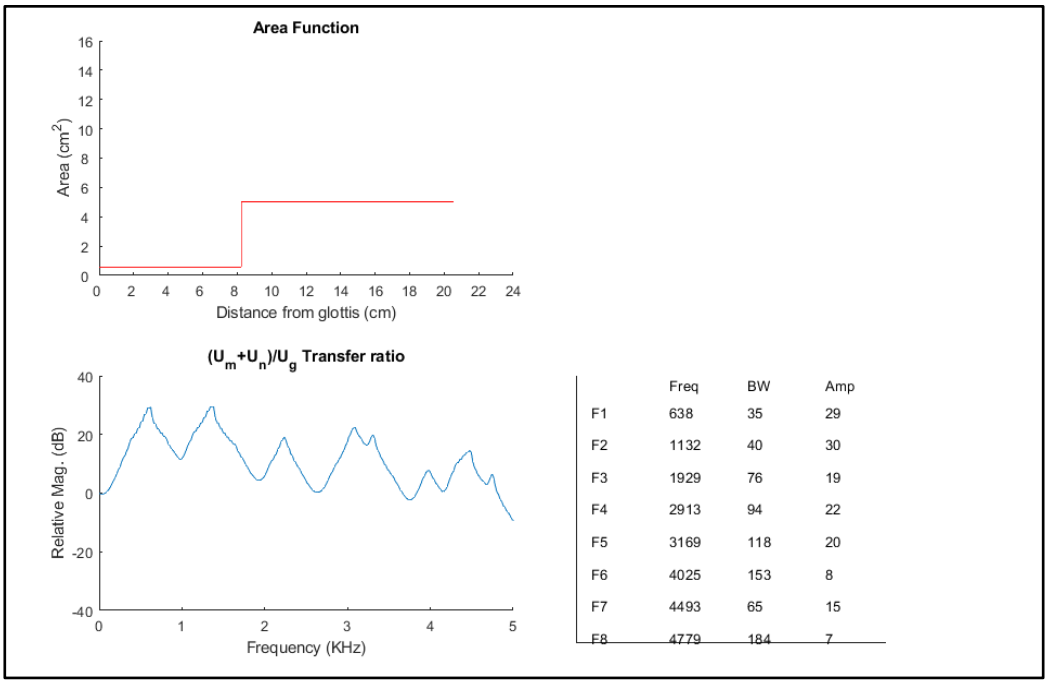
The results of the formant frequency analysis performed in VTCalcs are displayed in Table 13. The standard International Phonetic Alphabet (IPA) values for F1 and F2 frequencies are listed for reference (International Phonetic Alphabet, 1999). Note that the average heidelbergensis results are the VTCalcs formant frequency outputs of the AVG *heidelbergensis* VT dimensions from Table 12, not the mean of the hertz values of all *heidelbergensis* specimens.

Table 13

VTCalcs Formant Frequency Output						
	/a/ F1	**/a/ F2**	**/i/ F1**	**/i/ F2**	**/u/ F1**	**/u/ F2**
AVG *sapiens*	832	1548	304	2389	256	544
IPA Value	850	1610	240	2400	250	595
Arago	624	1386	183	1321	272	712
Atapuerca	663	1130	170	1432	264	608
Kabwe	686	1097	170	1498	264	617
Petralona	592	1050	155	1271	256	560
Steinheim	627	1067	160	1354	256	584
AVG *heidelbergensis*	638	1132	256	1374	272	616
AVG *heidelbergensis* No Arago	631	1125	256	1359	272	610

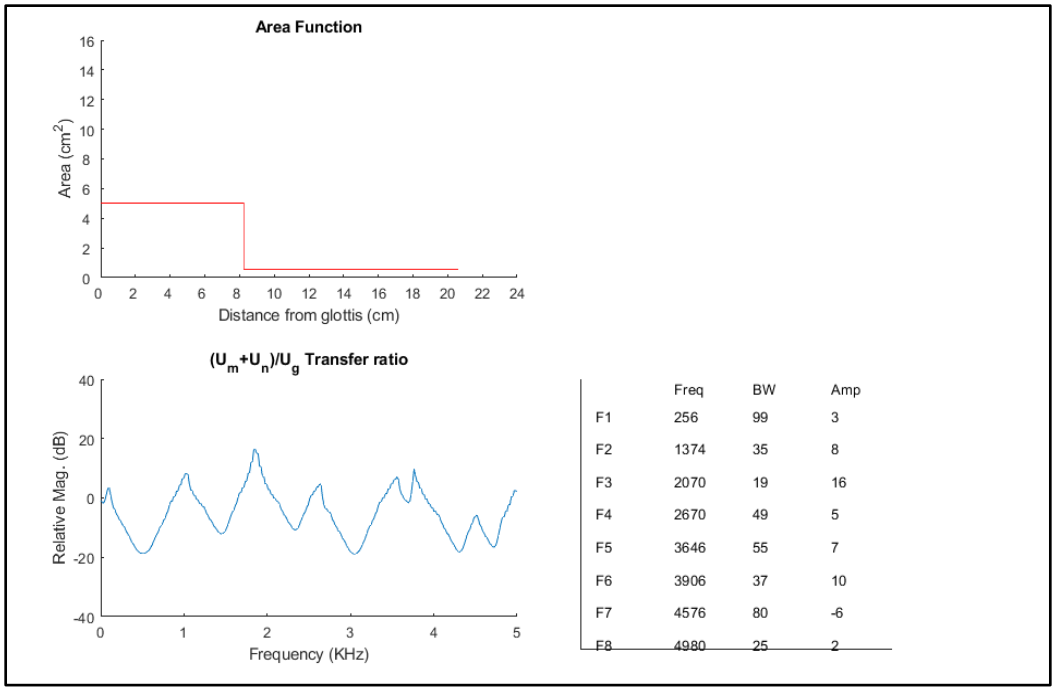
Values in Hz.

As with *Homo sapiens*, the *Homo heidelbergensis* /a/ is characterized by a high F1 and high F2, /i/ is characterized by a low F1 and high F2, and /u/ is characterized by a low F1 and low F2. See Figures 19, 20 and 21 for the detailed formant frequency outputs for the AVG *heidelbergensis* values.



	Freq	BW	Amp
F1	638	35	29
F2	1132	40	30
F3	1929	76	19
F4	2913	94	22
F5	3169	118	20
F6	4025	153	8
F7	4493	65	15
F8	4779	184	7

Figure 19. VTCalcs Formant Frequency Output for Average heidelbergensis Vocal Tract Dimensions - /a/



	Freq	BW	Amp
F1	256	99	3
F2	1374	35	8
F3	2070	19	16
F4	2670	49	5
F5	3646	55	7
F6	3906	37	10
F7	4576	80	-6
F8	4980	25	2

Figure 20. VTCalcs Formant Frequency Output for Average heidelbergensis Vocal Tract Dimensions - /i/

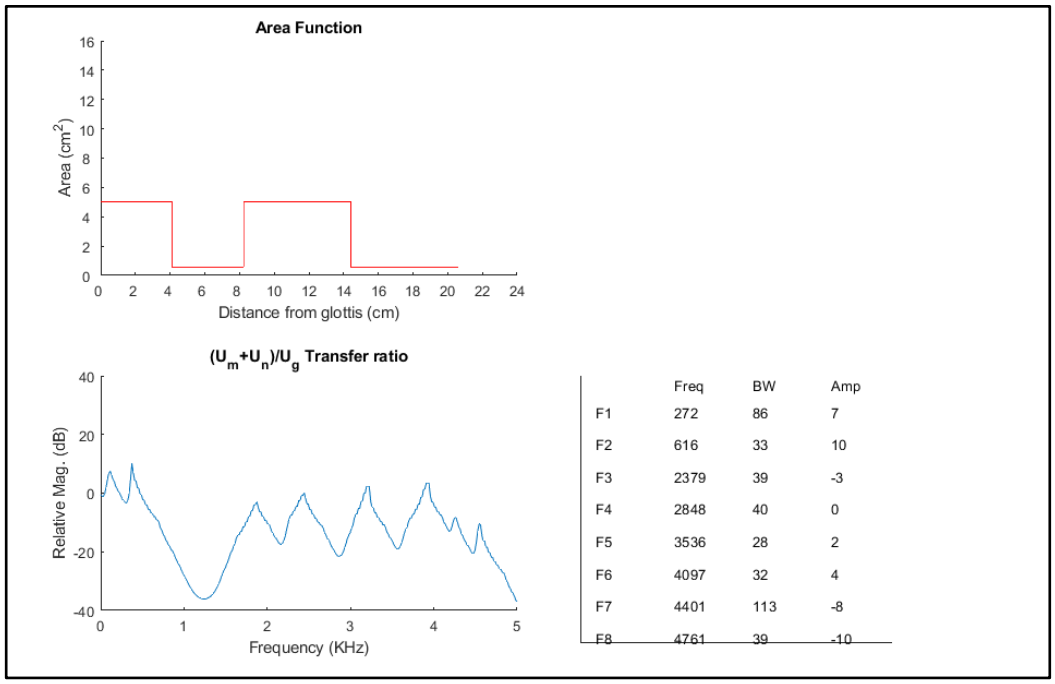


Figure 21. VTCalcs Formant Frequency Output for Average heidelbergensis Vocal Tract Dimensions - /u/

Figure 22 presents a visual depiction of the *Homo heidelbergensis* vowel space reconstructed through this research, compared to the locations of the *Homo sapiens* formant frequencies which were also gathered through this research. The *H. sapiens* formant values are the AVG *sapiens* values from Table 12, and the *H. heidelbergensis* formant values are the AVG *heidelbergensis* from Table 12. This reconstruction of the *H. heidelbergensis* vowel space indicates a notably smaller range of vowel sounds, the implications of which will be examined in the Discussion section. See Figure 23 for a comparison of these vowel spaces with the standard IPA values. See Figure 24 for the individual vowel space of each *H. heidelbergensis* specimen measured in this study.

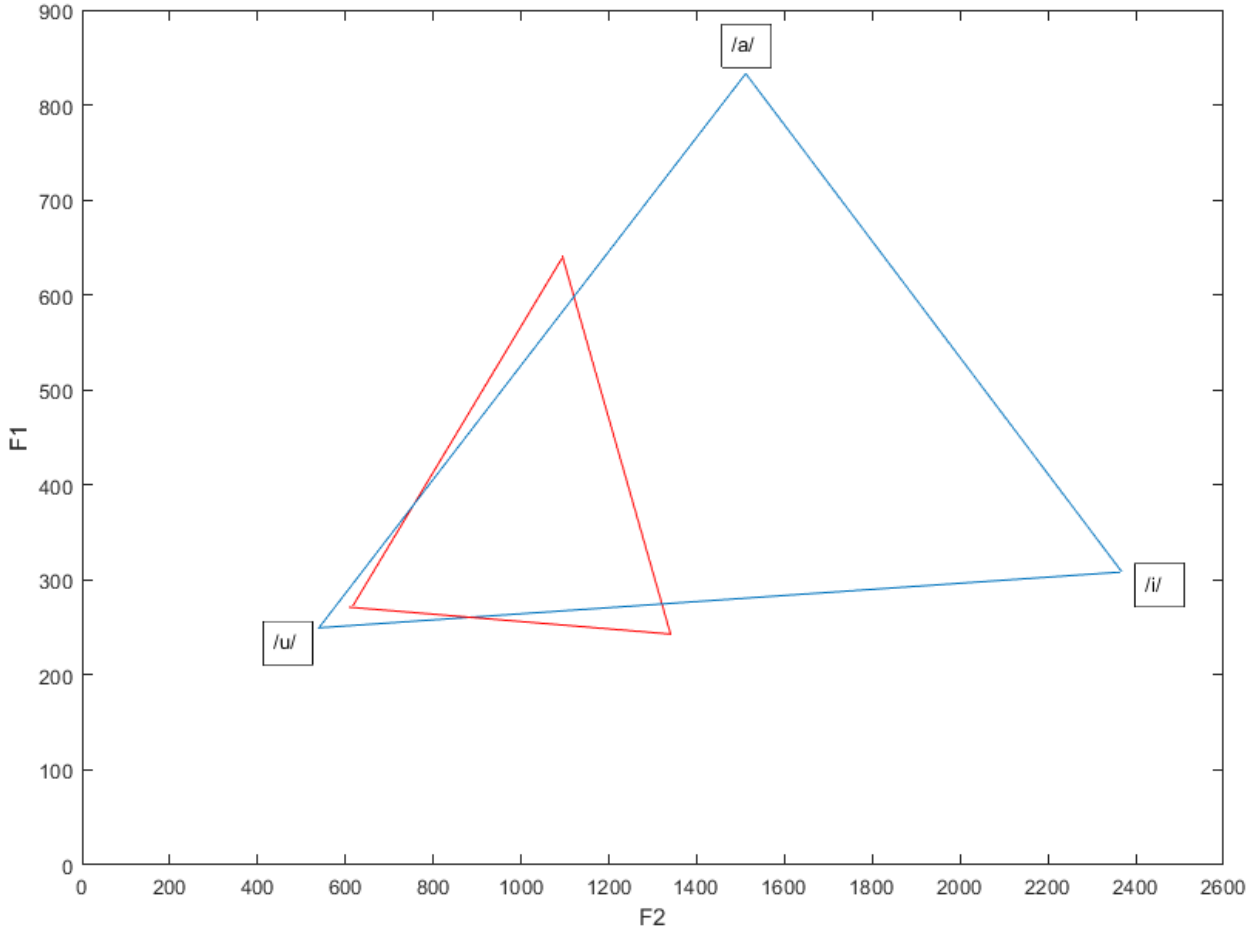


Figure 22. Representation of the Homo sapiens and Homo heidelbergensis Vowel Spaces

***Blue**: Homo sapiens. **Red**: Homo heidelbergensis. Values in Hz.*

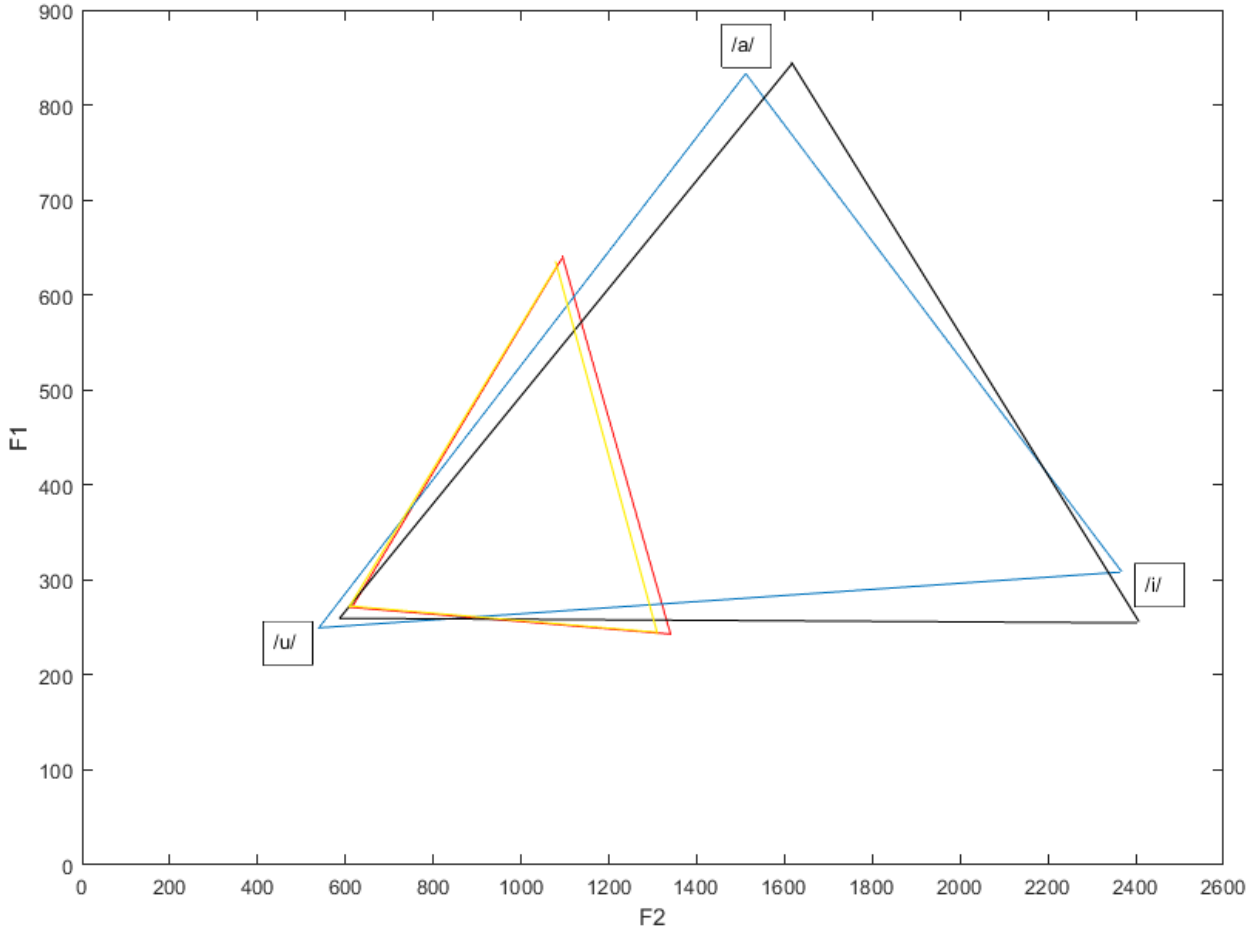


Figure 23. Representation of the Homo sapiens and Homo heidelbergensis Vowel Spaces

***Blue**: Homo sapiens. **Red**: Homo heidelbergensis. **Black**: IPA Values. **Yellow** Homo heidelbergensis (No Arago)*

Values in Hz.

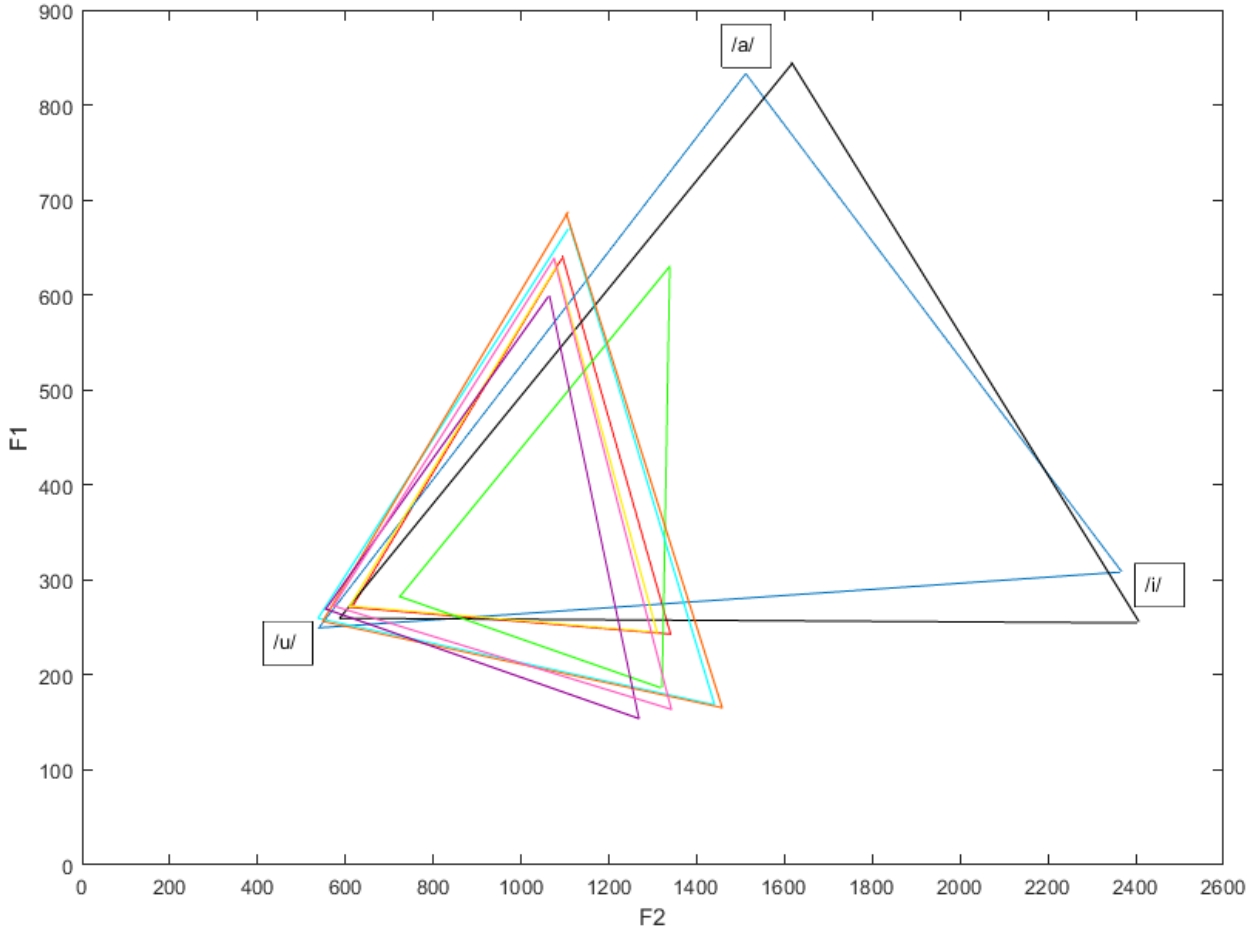


Figure 24. Representation of the Homo sapiens and Homo heidelbergensis Vowel Spaces

***Blue**: Homo sapiens. **Red**: Homo heidelbergensis. **Black**: IPA Values. **Yellow** Homo heidelbergensis (No Arago)*

***Green**: Arago 21. **Light Blue**: Atapuerca 5. **Orange**: Kabwe 1. **Purple**: Petralona 1. **Pink**: Steinheim*

Values in Hz.

CHAPTER 5: DISCUSSION

The results from the formant frequency analysis indicate that *Homo heidelbergensis* had a markedly smaller vowel space than modern humans. Previous research on hominin vowel space indicates that as the VTv:VTh ratio of a two-tube acoustic system approaches 1:1, the size of the vowel space increases (Boë, 2013; Barney, 2012; Granat, 2007). As such, the size of the vowel space can be traced in part by observing the changes in this ratio throughout human evolution. Boë et al. measured six adult *H. neanderthalensis* specimens and determined that this ratio was approximately 8:10 (2013). The measurements from the modern human sample in my research indicate a *H. sapiens* ratio of 9:10, and the *H. heidelbergensis* measurements indicate a ratio of 2:3. The majority of the shifts in overall VT length in these species occurred in the VTh. In other words, as overall prognathism and oral cavity length decreased throughout human evolution, the ratio of the VTv:VTh approached 1:1.

However, this research is not an adequate predictor of articulatory control or the mental capacity that may be needed to communicate through language. As previously discussed, the size of the hypoglossal canal does not adequately correlate to control over the tongue, and currently there are no conclusive methods to test for articulatory control of the tongue on fossil remains. If the mental capacity to communicate through language emerged before hominins had the ability to rapidly sequence vocal utterances, it is possible that the earliest forms of language may have resembled the protolanguage that Bickerton surmised (1990).

While it is known that cranial capacity increased throughout human evolution, it is not known at what point hominins would have the cognitive abilities to communicate through language. It was recently discovered that a mutation in the gene FOXP2 influences linguistic

capabilities (Enard, 2002). Half of the members of a modern human family dubbed the "KE Family" have a mutation in the FOXP2 gene which inhibits linguistic and grammatical abilities, and causes severe issues with articulation. In the 140 million years between the emergence of rodents and the emergence of hominins, only one change in this gene occurred. In the 6 million years between the emergence of hominins and the present day, two changes occurred. While the precise moment of origin for these mutations are unknown, it is known that both *Homo neanderthalensis* and *Homo sapiens* share the "modern human" variant of the FOXP2 gene. The timing of these mutations, along with its strong links to linguistic abilities, suggests that one or both of these mutations were selected for due to their communicative advantages.

Although research on the vowel space cannot directly indicate the presence of these elements of speech, analysing the vowel space of a species can provide information about the range of vowel sounds they would have been able to produce. A vast majority of modern human languages have the vowels /a/, /i/, and /u/ (Boë, 2013). While all three of these vowels may not be a necessity for human-like language to exist, it has been suggested that the widespread presence of these vowels is owed to their distinguishability (Lieberman, 1972). Thus, it is notable that the formant frequencies of the *H. heidelbergensis* /a/ and /i/ are both significantly closer to /u/ when compared to modern humans. This suggests that the three corners of the *H. heidelbergensis* vowel space could have been less easily distinguished to the listener, and that the overall range of distinct sounds that *H. heidelbergensis* would have been capable of would have been limited when compared to modern humans.

CONCLUSIONS

To summarize, the analysis of the formant frequency outputs of five *Homo heidelbergensis* specimens reveals that the vowel space of these specimens was smaller than a modern human sample. While this provides a foundation in understanding the potential vowel sounds this species could have produced, a larger scope of speech related adaptations must be considered before it is known whether *H. heidelbergensis* could have communicated through vocal language. While some of the characteristics of the vocal tract likely had an origin unrelated to speech, such as the descent of the larynx, other more recent adaptations could indicate selection for a linguistic purpose, such as the disappearance of air sacs and the enlargement of the thoracic vertebral canal. Research on human evolution is largely dictated by the availability of evidence in the fossil record, and unfortunately critical fossils for understanding the speech anatomy of *H. heidelbergensis* have been found only recently, or not at all. The discovery of two *H. heidelbergensis* hyoids verified the lack of air sacs in the species, however the scarcity of intact thoracic vertebrae impede examination into the precision of their breathing control. With the fossils that are currently available, it is still possible for research on the vocal capabilities *H. heidelbergensis* to continue. By doing so, we continue to gain insight on the possible origins of human language.

Study Limitations

Due to the scarcity of well-preserved fossil specimens, the sample size of *Homo heidelbergensis* individuals used in this study is small. Three fossil casts were used in this study, and as such measurements taken on them should be viewed as an approximation of the actual

measurements. Because the original fossil cranium of the Steinheim specimen was significantly impacted, the cast used in this study is a model of the estimated anatomy of the individual, rather than a one-to-one replica of the original fossil. The left side of the fossil was damaged particularly in the region of the left temporal and zygomatic bones. Taking measurements on any cast with predicted anatomy has its benefits and drawbacks. Even with the original fossil, the true anatomical dimensions of the original individual are unobtainable without creating estimations. By measuring this fossil cast, it ensures that the measured dimensions reflect what is estimated to be anatomically accurate.

The posterior portion of the palate on the Arago 21 3D model we have received from the Smithsonian Institute was damaged. As such, the posterior nasal spine was not fully present. Measurements taken of that region, as well as multiple linear regression predicted values concerning the Arago 21 model should be viewed with caution.

None of the ANOVAs from the regression equation results were significant. The high p-value for these models is likely due to the small sample size. However, there was no way to increase the sample size for this study, as all available *Homo heidelbergensis* fossil specimens that retained the anatomical features that needed to be measured were measured for this study. The R^2 values indicate that the prediction formulas do predict a substantial amount of variation, so the results of the multiple linear regressions were ultimately used for the frequency analysis.

In a more general sense, paleoanthropology studies are often limited by small sample sizes and limited access to specimens that are available, hence the use of teaching specimens in this investigation. Since teaching specimens are often copies of original specimens, there can be differences in quality based on casting techniques and how much of the original fossil was

missing. In some cases, anatomical experts model missing regions by estimating missing anatomy or mirroring good sides onto bad sides. Moreover, original specimens are sometimes warped by geological processes over thousands of years during the process of fossilization. Some additional specimens of *H. heidelbergensis* exist, such as the Bodo skull, but were missing the necessary anatomy to be included in the current study. Consequently, for all of these aforementioned reasons, measurements and results from this investigation must be viewed with caution. Additional studies should be carried out if new *H. heidelbergensis* specimens are found.

REFERENCES

An, N. (2014, January 31). Human Spine. Retrieved March 29, 2018, from https://grabcad.com/library/human-spine-1

Barney, A., Martelli, S., Serrurier, A., & Steele, J. (2012). Articulatory capacity of Neanderthals, a very recent and human-like fossil hominin. Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1585), 88-102.

Bickerton D. (1990). Language and species. Chicago: University of Chicago Press

Boë, L. J., Granat, J., Heim, J. L., Badin, P., Barbier, G., Captier, G., ... & Schwartz, J. L. (2013). Reconstructed fossil vocal tracts and the production of speech. New perspectives on the origins of language, 144, 75.

Boer, B. D. (2012). Loss of air sacs improved hominin speech abilities. Journal of Human Evolution, 62(1), 1-6.

Carretero, J. M., Lorenzo, C., & Arsuaga, J. L. (1999). Axial and appendicular skeleton of Homo antecessor. Journal of Human Evolution, 37(3-4), 459-499.

Cartmill, M. (1998). The gift of gab. Discover 19:56-64.

Corballis, M. C. (2003). From mouth to hand: gesture, speech, and the evolution of right-handedness. Behav Brain Sciences 26, 199-260.

Enard, W., Przeworski, M., Fisher, S. E., & Lai, C. S. (2002). Molecular Evolution of FOXP2, a Gene Involved in Speech and Language. Nature, 418, 869-871.

Fitch, W. T., & Reby, D. (2001). The Descent of the Larynx is Not Unique to Humans. The Royal Society, 268, 1669-1675.

Granat, J., Boë, L., Badin, P., Pochic, D., Heim, J., Peyre, E., & Benoit, R. (2007). Prediction of the Ability of Reconstituted Vocal Tracts of Fossils to Produce Speech. Saarbrücken, 6-10, 381-384.

International Phonetic Association. (1999). Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet. Cambridge University Press.

Johansson, S. (2006). Constraining The Time When Language Evolved. The Evolution of Language.

Jungers, W. L., Pokempner, A. A., Kay, R. F., & Cartmill, M. (2003). Hypoglossal Canal Size in Living Hominoids and the Evolution of Human Speech. Human Biology, 75(4), 473-484.

Kay, R.F., M. Cartmill, and M. Balow. (1998). The hypoglossal canal and the origin of human vocal behavior. Proc. NatL Acad. Sci. USA 95:5417-5419.

Leggett, C. (2014, October 13). Human Hyoid Bone Model for Printing. Retrieved March 29, 2018, from https://3dprint.nih.gov/discover/3dpx-000718

Lele, S. R., & Richtsmeier, J. T. (2001). An Invariant Approach to Statistical Analysis of Shapes (Chapman & Hall/CRC Interdisciplinary Statistics) (1st ed.). London: Chapman & Hall.

Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. Trends in cognitive sciences, 4(5), 187-196.

Lieberman, P., Crelin, E. S., & Klatt, D. H. (1972). Phonetic Ability and Related Anatomy of the Newborn and Adult Human, Neanderthal Man, and the Chimpanzee. American Anthropologist, 74(3).

MacLarnon, A. M., & Hewitt, G. P. (1999). The evolution of human speech: The role of enhanced breathing control. American Journal of Physical Anthropology, 109(3), 341-363.

Mallison, H. (2013). Photogrammetry tutorial 3: turntables. Retrieved from https://dinosaurpalaeo.wordpress.com/2013/12/20/photogrammetry-tutorial-3-turntables/

Martinez, I., Arsuaga, J. L., Quam, R., Carretero, J. M., Gracia, A., & Rodriguez, L. (2007). Human hyoid bones from the middle Pleistocene site of the Sima de los Huesos. Journal of Human Evolution.

Martínez, I., Rosa, M., Quam, R., Jarabo, P., Lorenzo, C., Bonmatí, A.,Arsuaga, J. (2012). Communicative capacities in Middle Pleistocene humans from the Sierra de Atapuerca in Spain. Quaternary International.

Moraes, C. A., Dias, P. E., & Melani, R. F. (2014). Demonstration of protocol for computer-aided forensic facial reconstruction with free software and photogrammetry. Journal of Research in Dentistry, 2(1), 77.

Nishimura, T., Mikami, A., Suzuki, J., & Matsuzawa, T. (2006). Descent of the hyoid in chimpanzees: evolution of face flattening and speech. Journal of Human Evolution, 51(3).

Rosas, A. (1997). A gradient of size and shape for the Atapuerca sample and Middle Pleistocene hominid variability. Journal of Human Evolution, 33(2-3).

Rightmire, G. P. (2008). Homo in the middle pleistocene: Hypodigms, variation, and species recognition. Evolutionary Anthropology: Issues, News, and Reviews,17(1), 8-21.

Samarin, W. J. (1972). Glossolalia. Psychology Today, 6, 48-50.

Steele, J., Clegg, M., & Martelli, S. (2013). Comparative Morphology of the Hominin and African Ape Hyoid Bone, a Possible Marker of the Evolution of Speech. Human Biology, 85(5).

Story, B. H., Titze, I. R., & Hoffman, E. A. (1996). Vocal tract area functions from magnetic resonance imaging. Journal of the Acoustical Society of America, 100, (1).

Tattersall, I. (2000). Paleoanthropology: The last half-century. Evolutionary Anthropology: Issues, News, and Reviews, 9(1), 2-16.

University of Southern California, Signal Analysis and Interpretation Laboratory. (n.d.). Formant Frequencies. Retrieved from http://sail.usc.edu/~lgoldste/General_Phonetics/Source_Filter/SFc.html

Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., Gentry, L. R. (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study. The Journal of the Acoustical Society of America, 125(3), 1666-1678.

Zhang, S., Block, K. T., & Frahm, J. (2009). Magnetic resonance imaging in real time: Advances using radial FLASH. Journal of Magnetic Resonance Imaging, 31(1), 101-109.